BIOMETRIKA

VOL. 47, PARTS 3 and 4 DECEMBER, 1960

CONTENTS

MEMOIRS:

LESLIE, P. H. and GOWER, J. C. The properties of a stochastic model for the predator-prey type of interaction between two species—BLOM, GUNNAR. Hierarchical birth and death processes I. Theory—BLOM, GUNNAR. Hierarchical birth and death processes II. Applications—GLENN, W. A. A comparison of the effectiveness of tournaments—PEARCE, S. C. Supplemented balance—FOLKS, JOHN LEROY and KEMPTHORNE, OSCAR. The efficiency of blocking in incomplete block designs—HAIGHT, FRANK A. Queueing with balking, II—DAVID, H. A. and PEREZ, CARMEN, A. On comparing different tests of the same hypothesis—FARLIE, D. J. G. The performance of some correlation coefficients for a general bivariate distribution—MCFADDEN, J. A. Two expansions for the quadrivariate normal integral—LAHA, R. G. and LUKACS, E. On a problem connected with quadratic regression—MCCULLOUGH, ROGER, S., GURLAND, JOHN and ROSENBERG, LLOYD. Small sample behaviour of certain tests of the hypothesis under variance heterogeneity—SUKHATME, BALKRISHNA V. Power of some two-sample non-parametric tests—EWAN, W. D. and KEMP, K. W. Sampling inspection of continuous processes with no autocorrelation between successive results—BLYTH, COLIN R. and HUTCHINSON, DAVID W. Table of Neyman-shortest unbiased confidence intervals for the binomial parameter—BENNETT, B. M. and HSU, P. On the power function of the exact test for the 2×2 contingency table—SIMPSON, J. A. and WELCH, B. L. Table of the bounds of the probability integral when the first four moments are given—SEVERO, NORMAN C. and ZELAN, MARVIN. Normal approximation to the chi-square and non-central F probability functions—BARTON, D. E., DAVID, F. N. and O'NEILL, ANNE F. Some properties of the distribution of the logarithm of non-central F—LINDLEY, D. V., EAST, D. A. and HAMILTON, P. A. Tables for making inferences about the variance of a normal distribution—BARTON, D. E., DAVID, F. N. and MERRINGTON, M. Tables for the solution of the exponential equation, $\exp(-a)+ka=1$—MISCELLANEA: Contributions by G. E. BARDWELL, B. R. BHAT and J. GANI, D. R. COX, R. N. CURNOW, N. GILBERT, W. M. HARPER and J. A. MACDONALD, N. L. JOHNSON and D. H. YOUNG, M. G. KENDALL, A. N. KSHIRSAGAR, H. LINHART, T. A. RAMASUBBAN—REVIEWS—OTHER BOOKS RECEIVED.

The subscription, payable in advance, is 54s (or $8) per volume (including postage). Cheques should be made payable to Biometrika, *crossed* "a/c Biometrica Trust" *and sent to the Secretary, Biometrika Office, University College, London, W.C.1. All foreign cheques must be drawn on a bank having a London agency.*

Issued by THE BIOMETRIKA OFFICE, University College, London

THE CARBORUNDUM COMPANY LIMITED

has a vacancy for a

Quality Control Statistician

in its Technical Branch at Trafford Park, Manchester. The successful applicant will be required to work on problems of applying statistical quality control methods in the manufacture of the company's various product lines. A good science or mathematics degree is essential. The post offers good prospects with ample scope for advancement. Good working conditions, bonus scheme and generous pension scheme. All applications, which will be treated in strict confidence, should include details of age, qualifications and experience and should be addressed to the Manager, Quality Control Department, Technical Branch, The Carborundum Company Limited, Trafford Park, Manchester, 17.

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part
20s.

Annual Subscription
Post Free £4 4s.
U.S. $12

Principal Contents of Volume 123 Part 4 1960

The Relationship between Family Size and Family Expenditure. By F. G. Forsyth (With Discussion).

An Experimental Investigation of the Robustness of Certain Procedures for Estimating Means and Regression Coefficients. By D. G. Champernowne.

Statistical Problems connected with the 1961 Population Census. By B. Benjamin.

Output and Productivity Changes in British Manufacturing Industry 1948-1954. By R. J. Nicholson and S. Gupta.

Annual Report of the Council.
Proceedings of the 126th Annual General Meeting.

Reviews

Current Notes.

Obituary:
Georges Darmois
Oskar Anderson

Statistical and Economic articles in recent periodicals.

Additions to Library.

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638

Economic and Social Statistics

The Department of Technical Co-operation would like to be in touch with well qualified men and women with experience in fields such as economic statistics including national income estimates and family expenditure surveys who would welcome an assignment in under-developed countries. The Department receives a number of requests for Technical Assistance in this form from the U.K. Details of specific appointments will be discussed with suitably qualified applicants and should a definite offer be in prospect, the Department would be prepared to approach the present employer of the applicant on the question of leave of absence.

A letter giving brief details of experience and main interests should be sent in the first place to the Director-General, Department of Technical Co-operation, Sanctuary Buildings, Great Smith Street, London, S.W.1. quoting reference BC 343/01.

One-Year International Post Graduate Course In Operational Research

(Sponsored by O.E.C.D.)

The next Course in Operational Research for Engineers, Scientists and Industrialists of Degree standard will begin in October, 1962. A thorough training will be given, not only in the theoretical techniques, but also in the application of these techniques to practical problems in Operational Research.

The Course is recognised by O.E.C.D. as the European Centre for training in Operational Research. Students sponsored by organisations in O.E.C.D. countries will receive grants. For details and application forms, please apply to:

The Registrar,
The College of Aeronautics,
Cranfield,
Bletchley, Bucks.,
England.

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

Ann. Sub. incl. postage
£3 2s. U.S. $9

Contents of Volume 24 Part 1 1962

Some Results on Inventory Problems (with Discussion) By *A. R. Thatcher*

Efficient Estimates and Optimum Inference Procedures in Large Samples (with Discussion) By *C. R. Rao*

A Waiting Line with Interrupted Service, including Priorities By *D. P. Gaver*

Time Dependent Solution of the 'Head-of-the-Line' Priority Queue By *N. K. Jaiswal*

Inequalities for First Emptiness Probabilities of a Dam with Ordered Inputs By *J. Gani and B. Pyke*

Congestion Systems with Incomplete Service By *F. Downton*

The Estimation of Slope when the Errors are Auto-correlated By *G. M. Jenkins and J. Chanmugan*

The Analysis of Stationary Processes with Mixed Spectra By *M. B. Priestley*

The Estimation of Spectral Densities By *H. E. Daniels*

Subsidiary Sequences for Solving Lesser's Least-squares Graduation Equations By *A. W. Joseph*

Some Examples of Bayes' Method of the Experimental Determination of Probabilities *a priori* By *Sir Ronald Fisher*

Exact Fiducial Limits in Non-linear Estimation By *E. J. Williams*

On the Effect of Non-normality on the Estimation of Components of Variance By *M. Atiqullah*

On the Distribution of Certain Quadratic Forms in Normal Variates. By *B. R. Bhat*

A Note on a Generalised Inverse of a Matrix with Applications to Problems in Mathematical Statistics By *C. R. Rao*

Large Sample Restricted Parametric Tests By *John Aitchison*

On Multivariate Sign Tests By *B. M. Bennett*

A Note on Interactions in Contingency Tables By *R. L. Plackett*

Interactions in Multi-factor Contingency Tables By *J. N. Darroch*

On the Asymptotic Sufficiency of Certain Order Statistics By *M. S. Rohatgi*

A New Asymptotic Expansion for the Normal Probability Interval By *H. Ruben*

Estimators for the Product of Arithmetic Means By *G. J. Glasser*

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638

CONTENTS (*continued*) *Page*

No. 3. NOVEMBER

The Household Expenditure Enquiry of the Ministry of Labour. Variability in the 1953-54 Enquiry .. .. *W. F. F. Kemsley* 117

Inter-week Variations in Expenditure recorded during a Two-week Survey of Family Expenditure .. .. *Robert Turner* 136

The Variable Lot-size Acceptance Sampling Plan for Continuous Production .. .. .. *D. R. Read and D. W. Beattie* 147

Empirical Evidence on Optimum Motor Taxes for the United Kingdom .. .. *A. A. Walters* 157

On the Price Consciousness of Consumers *André Gabor and C. W. J. Granger* 170

Book Reviews:

Stochastic Population Models in Ecology and Epidemiology M. S. Bartlett 189

Statistical Theory and Methodology in Science and Engineering K. A. Brownlee 189

Sample Design in Business Research, W. Edwards Deming 190

Tables of the Hypergeometric Probability Distribution Gerald J. Lieberman and Donald B. Owen 191

CONTENTS

No. 1. MARCH

		Page
Some Industrial Applications of Multivariate Analysis	*B. A. M. Thomas*	1
Measurements and Standards	*Freda Conway*	9
Some Notes on the Precision of the Gradient of an Estimated Straight Line	*Donald G. Beech*	14
An Introduction to Dynamic Programming ..	*M. G. Simpson*	32
Dynamic Programming and Decision Theory	*D. V. Lindley*	39
A Method of Allowing for 'Not-at-home' Bias in Sample Surveys	*D. J. Bartholomew*	52
Notes and Comments		60
Meetings of Sections and Local Groups of the Royal Statistical Society		61
Book Reviews:		
Modern Probability Theory and Its Applications, Emanuel Parzen ..		63
Quality Control, Norbert L. Enrick		64

No. 2. JUNE

A Critical Examination of Simon's Model of Certain Distribution Functions in Linguistics	*Gustav Herdan*	65
An Application of Non-Parametric Statistics to Truncated Selection	*G. M. Tallis*	77
Confounding in 3^3 Factorial Experiments in Nine Blocks	*D. Wilkie*	83
Interviewer Differences among Automobile Purchasers	*Allen F. Jung*	93
The Analysis of a $3^2 \times 2^2$ Factorial Experiment with Confounding	*R. H. E. Inkson*	98
Notes and Comments		108
Meetings of Sections and Local Groups of the Royal Statistical Society		110
Book Reviews:		
Linear Programming in Industry—Theory and Applications: An Introduction, Sven Danø		112
A Second Course in Statistics, Robert Loveday		112
The Chosen Few: An Examination of Some Aspects of University Selection in Britain, W. D. Furneaux		113
Publications Received		116

Digitized by the Internet Archive
in 2025

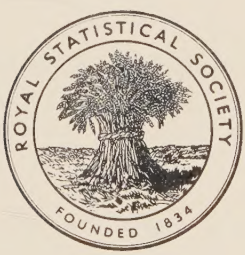
https://archive.org/details/journal-of-the-royal-statistical-society-series-c_1961_10

APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITOR
DONALD G. BEECH

ASSISTANT EDITOR
LESLIE A. DICKS-MIREAUX



VOLUME TEN

Oliver and Boyd Ltd.
London: 39A Welbeck Street, W.1
Edinburgh: Tweeddale Court

1961

← Table of Contents
behind Title Page

APPLIED STATISTICS

SOME INDUSTRIAL APPLICATIONS OF MULTIVARIATE ANALYSIS

B. A. M. THOMAS*

In the past, multivariate techniques have been applied mainly in psychological studies, but it has already been suggested in this journal that this need not preclude other applications. Here we have some industrial examples to support the suggestion.

Introduction

It is a curious thing that whereas in recent years there has been a great upsurge in the use of multiple regression techniques in industry, no corresponding increase has been recorded in the use of multivariate analysis. This is all the more remarkable since electronic computers, which are largely responsible for the increased popularity of multiple regression analysis, allow multivariate techniques such as principal component analysis to be used as easily as regression analysis.

The reason is probably that industrial situations involving many variables usually resolve themselves into a search for definite relationships to enable some of the variables to be predicted from the remainder. Such problems obviously demand the use of regression techniques or something similar, in order to provide the sort of precise answers required. The types of results given by multivariate analysis, on the other hand, seem much more vague and less obviously useful.

The differences between the two sets of techniques may be likened to the differences between a map of a town and photographs of views of the town. The map, like regression analysis, can provide precise answers to well-defined questions. The kind of question which might be asked is 'How can one get from A to B?' The photographs cannot answer this kind of question very well but are still useful in interpreting the information given by the map and in deciding whether the journey from A to B is worth making.

The great problems in regression analysis, in fact, are deciding which regression analyses should be performed and interpreting the results. It would appear that one of the most useful fields for multivariate analysis in industry is in helping to solve these problems. The way in which this can be done must vary from problem to problem, but some indications may be obtained from the two practical examples which follow.

* This article is based on a paper by the late Mr B. A. M. Thomas, formerly of the British Petroleum Company Limited, which was read at the Durham Conference of the Royal Statistical Society, September 1960.

Specification Tests on Petroleum Products

Tests Required. Increasingly stringent quality requirements for petroleum products have led to a great profusion of specification tests applicable to each product. It is not unusual, for example, for upwards of twenty different tests to be carried out on each batch of material leaving a refinery. Many of these tests are of long standing and are of a highly empirical nature so that it is often exceedingly difficult to discern what relationships, if any, exist between the various tests. This can sometimes make the task of meeting the specifications unnecessarily difficult.

Study of Gas Oils. This situation is a fairly typical one in which multivariate techniques can usefully be applied. Accordingly a series of studies of various products was undertaken at the British Petroleum Research Centre. The first of these, concerned with gas oils, illustrates the sort of results which can be obtained.

Gas oil is a petroleum product intermediate in boiling range between kerosene (paraffin) and fuel oil. At one time it was used for making gas to enrich water-gas, hence its name, but it is now used mainly as a fuel for high-speed diesel engines and for domestic heating installations. It is manufactured in a number of grades, to different specifications, and is normally subjected to a large number of tests. The tests required for one specification may not necessarily be the same as for another. However, a search through the available data permitted about eighty sets of results to be selected which contained a common set of twelve tests. These results had been obtained over the course of several years on gas oils, manufactured in all parts of the world, which had been subjected to a number of different refinery processes. The tests considered are set out in Table I.

TABLE I

Variables used in analysis

Specific gravity
Distillation 10% point
Distillation 50% point
Distillation 90% point
Flash point
Sulphur content
Kinematic viscosity
Cloud point
Pour point
Carbon residue
Aniline point
Diesel index

Principal Components Analysis. The first step of the investigation was to form the matrix of sums of squares and products of the deviations of the variables and thence to obtain the covariance matrix. Then, since a principal components analysis was envisaged, the correlation matrix was formed as a rough and ready means of avoiding the scaling problem associated with the technique.

The purpose of the analysis is to account for the variation and intercorrelation of the properties in terms of a number of orthogonal com-

ponents, these components being linear functions of the properties. They are characterised by their eigenvalues, the characteristic roots of the correlation matrix, which indicate their relative importance, and their associated eigenvectors, which give the coefficients in the linear functions. The eigenvalues and eigenvectors were found by a straightforward, even if lengthy, iterative process. The eigenvalues obtained are given in Table II, and the vectors corresponding to the first five of these roots are given in Table III.

TABLE II

Eigenvalues of gas oil matrix

Eigenvalues	Cumulative Sums
4·76	4·76
2·95	7·71
1·54	9·25
1·16	10·41
0·80	11·21
0·28	11·49
0·24	11·73
0·14	11·87
0·05	11·92
0·05	11·97
0·02	11·99
0·01	12·00

TABLE III

Eigenvectors of gas oil matrix

Variable	1	2	3	4	5
Specific gravity	0·88	−0·31	−0·74	−0·24	−0·11
Distillation 10% point	0·59	0·98	−0·45	−0·01	−0·07
Distillation 50% point	0·95	0·41	0·10	−0·20	−0·01
Distillation 90% point	0·57	−0·83	0·77	0·02	−0·21
Flash point	0·54	0·82	−0·50	0·37	0·01
Sulphur content	0·22	0·09	0·30	1·00	−0·32
Kinematic viscosity	1·00	0·32	0·12	−0·14	−0·11
Cloud point	0·63	−0·60	0·90	−0·10	−0·11
Pour point	0·88	−0·16	0·46	0·13	0·18
Carbon residue	0·44	−0·43	−0·16	0·28	1·00
Aniline point	0·14	1·00	0·93	−0·25	0·18
Diesel index	−0·56	0·78	1·00	0·03	0·18

Interpretation of Results. Although no tests have been applied to ascertain the dimensionality of the system it is apparent that the first five roots account for a very large part of the total variation. The rather heuristic method of expressing the sum of the largest roots as a proportion of the total indicates that these five roots account for about 93% of the variation. If the precisions of the various specification tests are borne in mind, it is reasonable to suppose that much of the residual variability indicated by the seven smallest roots is experimental error.

In this case the vectors corresponding to the five largest roots appear to have some intrinsic interest. The first of the vectors has large positive coefficients corresponding to viscosity, specific gravity, 50% distillation

point, and pour point. These variables all give an indication of the position of a particular gas oil fraction in the crude oil from which it was drawn. High values for each of these variables are obtained with gas oils drawn from the fuel oil end of the range and low values with those from the kerosene end. This vector can thus be thought of as measuring the position of the fraction.

The second vector has large positive coefficients in the variables concerned with the low boiling end of the oil, such as 10% distillation point and flash point, and negative coefficients in the variables, such as 90% distillation point, concerned with the high boiling end. This vector, then, indicates the narrowness of the fraction.

Similarly the third vector is associated with the proportions of various hydrocarbon types in the gas oil. The fourth and fifth vectors appear to be straightforward measures of the sulphur content and the carbon residue of the oil.

Implications of the Analysis. Although twelve test variables were included in the analysis, the space spanned by the results was effectively only five-dimensional. This implies a considerable amount of redundancy in the original set of twelve tests. In normal practice little use is made of this redundancy and it is plain that considerable advantages would accrue from either reducing it or making more effective use of it.

The most obvious savings would result from reducing the redundancy by eliminating some of the specification tests. By this means the amount of testing effort required could in theory be reduced by about seven twelfths. In practice it would take some years and a considerable amount of negotiation with the consumers to obtain acceptance of such a reduction in testing.

A more easily realisable benefit could be obtained by using the redundancy to obtain more precise values for the individual test results. Each of the twelve test variables can be estimated by a regression equation from observations of the remaining eleven variables. The estimates can then be combined with the observed values to improve their precision. With some of the tests in use quite a worthwhile improvement in precision is obtainable in this way.

The use to which redundancy in the data is normally put is to provide the supervisors in the testing laboratory with a subjective assessment of the consistency of each set of results in order to detect any gross errors. There are several ways in which this assessment can be made more objective and more sensitive. An obvious one is to use Hotelling's T^2 to test for departures from the normal pattern of variation. This has in fact been done very effectively, but may be criticised on the grounds that it is a test for unusual properties of the oil as much as for an inconsistent set of results. If this distinction is important in any application, it can be taken account of by a minor modification of the T^2 test. By rotation of the axes of reference to correspond with the eigenvectors, T^2 can be split into orthogonal components corresponding to the various eigenvalues. The components which correspond to the five principal

roots would then indicate real changes in the properties of the oil. The sum of the remaining components forms the modified T^2 which, with the appropriate reduction in degrees of freedom, is used to test the consistency of the data.

If either the straightforward or the modified T^2 test indicates an important inconsistency in a set of results, the incorrect result can be pinpointed by a further test. This consists in comparing each observed result with the estimate obtained by means of a multiple regression function of the remaining results. These differences between observed and estimated results can conveniently be incorporated in control charts giving ready identification of the outlying results.

Once calculated, these differences can be put to a further use. It is common practice for petroleum testing laboratories to run continuous schemes to monitor the precision of the various methods of test. One form taken by such schemes is for a proportion of the routine samples entering the laboratory to be returned for re-testing after a suitable interval has elapsed. A record of the results obtained in such a scheme allows the within-laboratory standard deviation of each test to be checked. To obtain adequate coverage of all the tests carried out in a laboratory by this means requires an appreciable addition to the total work load of the laboratory in the form of repeat tests. The differences between observations and estimates used to test the self-consistency of sets of data can, however, be used to provide internal checks of precision without any need for repeat tests. Given a number of sets of results, the sum of squares of the differences between observations and estimates for each test method is a linear function of the variances of the twelve methods. Since there are twelve such sums of squares, one for each method of test, the matrix of coefficients can be inverted, leading to estimates of each variance. These estimates are biased by any non-linearities in the regression functions, but if the procedure is carried out continuously it can nevertheless provide indications of any changes in precision.

These are a few of the results which arise from the fact that the twelve tests on gas oils effectively measure only five characteristics. In the long run there are others which occur: the simplification of chemical engineering design calculations; the provision of firm foundations for the correlation of laboratory tests with field performance; and the more realistic setting of specifications.

Quality Control

Quality Control of Crude Oil Distillation Units. In most quality control applications it is customary to consider only one characteristic at a time and to control each property of a product separately. In the oil industry, and possibly in others, this is not always easy to do since the various properties to be controlled and the process variables which control them are often interrelated in quite complex ways. Consider, for example, a crude oil distillation unit. Crude oil is passed first

through a primary distillation column to remove the most volatile products, and then through the main atmospheric column to separate it into a range of products from gasolines down through kerosenes and gas oils to the atmospheric residue, which goes to feed a vacuum distillation column. A typical atmospheric column might have four products taken from the side streams.

Each of these products needs to have at least its specific gravity controlled at this stage. The process variables used to control them might be the amounts of product taken from each side stream and the temperature of the feed to the column. The specific gravities are correlated with one another and a change in any of the process variables changes the specific gravity of every product to a certain extent. In these circumstances the control of the distillation unit depends very much on the skill and experience of the operators. Recently an investigation was begun to discover whether control could be improved by the use of statistical methods.

Principal Components Analysis. Daily operating data relating to several months' operation of a given unit were collected. Twenty variables in all were chosen for study. These included properties of the crude oil, properties of the products, and values of the various process variables. Since only six process variables were used to control the unit in question it seemed desirable to check first of all that the dimensionality of the system of variables was not greater than six. To do this a principal components analysis of the correlation matrix of the data was carried out. The dimensionality was checked without need for testing, since after the first six eigenvalues the roots became trivially small. In this case, unlike others, it was not possible to attach meaningful labels to the vectors which were found. This was possibly due to the heterogeneous nature of the set of variables included in the analysis. However, because of this it seemed that attention would be better directed to techniques other than principal components analysis.

Further Work. This work is still proceeding, but it might be interesting to review some of the techniques appropriate to the problem. In this case the variables fall naturally into two groups: those describing product quality and the process variables. In such circumstances the use of Hotelling's canonical correlations might possibly reveal something of the structure of the system. However, more fruitful practical results might be expected from the application of multiple regression analysis to estimate the requisite changes in process variables as functions of deviations of product quality from the specification values. This analysis could follow the approach of Williams[1] to simultaneous regression problems.

Practical Difficulties

Scaling. In principal components analysis the results obtained are greatly dependent on the choice of scale for the measured variables. Where there is some natural scale which can apply equally to all

variables, as in some anthropometric work, this problem resolves itself. In other cases the solution commonly adopted is to standardise the scales to give a constant total variance for each variable, i.e. to work with the correlation matrix instead of the covariance matrix. This was the method adopted in the work described above, but is not a completely satisfactory solution. If there is adequate information concerning the variability of the methods of test used to obtain the data, a better solution might be to scale the results so as to give a constant precision of measurement for each scaled variable. The effect of this would be to give a more realistic weighting to the variables, taking into account the relative sensitivity of the test methods used to obtain them.

Identification of Components. The principal components found in an analysis depend quite obviously on the choice of variables included in the data. This implies that a set of components identified in an analysis, to which meaningful and apparently fundamental labels have been attached, can be changed beyond recognition by the addition or deletion of a few variables. This is because the principal components analysis shows the relative importance of factors only to the extent that this is reflected in the choice of variables included in the analysis and not in any absolute sense. For example, in the case of the gas oil study, a further analysis was made after eliminating five of the variables. The dimensionality of the space spanned by the remaining seven variables was still five but the five eigenvectors had changed completely. Some were still recognisable but were changed in relative magnitude; others appeared to be completely different. This was a clear demonstration that the apparently fundamental variables revealed by the analysis were not so much properties of the gas oil as of the particular combination of twelve tests used for the analysis. In this case the results were still of great value, but only because the tests included in the study were a reasonable reflection of the relative importance attached by the industry to the various properties of the oil. In other examples this might not necessarily be true.

Significance Testing. In the examples so far considered, the sizes of the eigenvalues were such that there was a fairly clear demarcation between the few large roots and the residue of comparatively small ones. This allowed the analyses to proceed without proper significance tests for the dimensionality of the systems. In fact it is difficult to discover in the literature any tests which would have been applicable to the situations which have been discussed. There is, however, a particular case for which published tests are available. This occurs if the measuring errors of the variables are all known so that the variables can be scaled to make the measuring errors equal, as was suggested in the paragraph on scaling. Tests applicable in this situation are given in reference 2. Alternatively, in the slightly more general case where only the relative sizes of the measuring errors are known, but their absolute values are not, use can be made of the technique of half-normal plots used by Daniel[3] for the analysis of factorial experimental contrasts. If there is

a reasonable number of observations this test is probably accurate enough for most practical purposes.

Conclusions

The techniques mentioned in this paper are not the only ones which might have been applied to the examples quoted. Moreover, the same conclusions might well have been reached without using multivariate analysis at all. However, I feel that by using multivariate techniques, in particular principal components analysis, the results have been obtained more rapidly than they would otherwise, and a greater insight has been gained into the nature of the problems.

Acknowledgements

I should like to thank my former colleagues at the British Petroleum Research Centre for their part in solving the problems and the Chairman and Directors of The British Petroleum Company for their permission to publish this paper.

REFERENCES

[1] Williams, E. J. (1958). 'Simultaneous regression equations in experimentation', *Biometrika*, **45**, 96.

[2] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, p. 329. John Wiley, New York.

[3] Daniel, C. (1959). 'Use of half-normal plots in interpreting factorial two-level experiments', *Technometrics*, **1**, 311.

MEASUREMENTS AND STANDARDS

FREDA CONWAY
University of Leicester

Social statisticians may find it difficult to devise statistics which will prove satisfactory to both administrators and social scientists unless they are willing to recognise and define the place of value judgements in social statistics.

Social statistics are frequently required to fulfil a double role. Both social scientists and administrators use them to measure and describe social phenomena, but administrators also use them to provide criteria or standards which can be employed in the determination of social policy. The use of statistics as standards implies the acceptance of certain social values, whereas their use as measurements does not.

The relationship between measurements and standards is not necessarily a simple one. In some cases standards must be adopted before a measurement can be made, but in other cases measurements that have become available are then used as standards. The measurement of poverty provides an example of the former case; the use of mortality rates in public health administration illustrates the latter.

The measurement of poverty involves the prior determination of a poverty line. In this country the first attempt to measure the extent of poverty was made by Charles Booth in his survey of London, which was begun in 1886. He was content to use a somewhat arbitrary poverty line—18 to 21 shillings per week for a moderate family—though he explained that this standard was based on his own knowledge of living conditions in the East End of London.

This survey was followed in 1899 by Seebohm Rowntree's first survey of York. Rowntree adopted a more scientific approach to the determination of a poverty standard, and later workers have mainly followed Rowntree's example.

Rowntree began by enumerating the things that are essential, i.e. food, housing, clothing, fuel, and light. He then estimated the minimum cost of providing for these needs and compared this cost with the income at the disposal of each family. His methods satisfied the definition of poverty given in the report of the Merseyside survey nearly thirty years later: 'A family is deemed to be living in poverty if the joint income of its members, supposing it to be wisely spent, is inadequate to buy for them the necessaries of life, taking into account the age and sex of the members of the family.' But before such a definition can be expressed numerically it must be interpreted in terms of prevailing ideas and standards. Conditions had changed in those thirty years: the standards of the Merseyside Survey were not exactly the same as those of the first

survey of York. One small example may be quoted since it is significant of changed attitudes. Scientific methods are most easily applied to food requirements, since much is known concerning the nutritional needs of human beings, but custom and convention play an important part in determining the kinds of foods that are eaten. Rowntree excluded butchers' meat from the minimum diet; later workers included it, since the Englishman is not traditionally a vegetarian.

More recently Mr Peter Townsend has suggested a new standard of poverty: 'The problem of whether or not a family is in poverty is best decided by finding whether its expenditure save for one or two involuntary overheads, such as rent and compulsory insurances, is less than that which actually secures minimum nutrition for a large number of working class families.' He suggests that such a standard would 'obviate the need for subjective decisions about the sums of money required for clothing, fuel and light and so on', though a 'subjective element remains since it is still necessary to choose the proportion of working class households whose members have an adequate diet and whose spending is to be considered in fixing the standard'. But is not this essentially the method adopted by Charles Booth? The difference lies in the fact that the modern social scientist can express his knowledge of living conditions in more exact terms than was possible at the time that Booth's survey was undertaken. This increase in exactness is of course a great gain; nevertheless it is still true that the concept of poverty implies a distinction between what is and what ought to be, and that the social scientist in accepting and using any measurement of poverty is accepting a standard of what ought to be, based on his interpretation of currently held values.

Not all social standards are obtained by direct consideration of the problem as in this case. Sometimes the measurement of a particular phenomenon makes comparisons possible and so provides a numerical standard which can be justified.

Mortality rates are simple measurements but public health authorities have consistently used the average rates for healthy groups as standards by which to judge the need for improvement among the less healthy. The use of mortality rates in this way hardly needs explicit justification, since the removal of the causes of premature death is a generally acceptable social policy. In most cases, however, it is necessary to justify the use of a measurement as a standard.

Now the social scientist who regards value judgements as outside his province may argue that the role of the social scientist is to produce a scale of measurement, such as that provided by mortality rates, leaving the administrator to determine the point on the scale which should be adopted as a standard. This point of view appears to be illustrated by a report published by a United Nations committee of experts—*International Definition and Measurement of Standards and Levels of Living*. This report suggests that a distinction should be made between indices of levels of living which could be used to describe the actual living condi-

tions of a people, and measurements of standards of living which might be defined as the level to which people aspired or to which they thought themselves entitled. Under such a system a poverty standard might be regarded as a point (or set of points) on a scale (or scales) measuring levels of living.

Unfortunately such a division of labour between the social scientist and the administrator is unlikely to prove practicable. Standards may be required long before the appropriate scales of measurement have been devised. Poverty standards have been used in social surveys for more than seventy years, but the discussion of measurements of levels of living is still in the theoretical stage. Also the choice of scale virtually determines the kind of standard that is adopted, since only those factors which have been considered in the construction of the measurement are automatically taken into account in fixing the standard.

The Census of Population for England and Wales has always provided some information concerning housing conditions and for many years the Registrar-General used the proportion of persons living at a density of more than two persons per room as a comparative index of overcrowding. The index was satisfactory for this limited purpose; but the use of this particular density in this way led to its being accepted, or at least attacked, as a standard of overcrowding, though it should have been obvious that a simple density measurement which takes no account of household composition could not be used to judge the adequacy of the housing accommodation in any area.

Although there is no one generally acceptable standard of overcrowding, all such standards agree that the household should have sufficient bedrooms to provide adequate separation of the sexes among persons above a minimum age, usually fixed at ten years. The standards differ in the allowance made for sleeping accommodation for younger children, the number of persons of the same sex permitted to share a bedroom, and the extent to which they recognise the need for living-rooms that are not used as bedrooms. The description of housing conditions given in the 1951 *Census Report* includes an analysis of persons-per-room densities by type of family and combines these to give weighted averages for different areas. In this case the relationship between these measurements and a number of overcrowding standards is also discussed; but even without this discussion, the new analysis would be more useful than the old, since it takes account of some of the factors that are used in determining standards of overcrowding.

The problems discussed here are relatively simple. So long as social reform is concerned to remedy obvious social evils, e.g. poverty, premature death, and overcrowding, there can be little doubt about what is good and desirable. In these circumstances it would seem essential for the social statistician to take the relevant social values into account in devising the measurements which will also be needed as standards.

But these relatively simple problems are not typical of those with which the majority of administrators and social scientists are now

concerned. Today social reform has more positive aims and there is more room for disagreement concerning the form of society which is desirable. At the same time statistical indices are being devised to measure various aspects of relatively complex situations, e.g. there are indices of industrial diversity and social mobility. The dangers involved in using some of these newer indices as standards are aggravated by the tendency of administrators and reforming citizens to think in terms of national averages, maxima, and minima.

Town-planners insist on the need to develop towns with a well-balanced industrial structure, a concept which presumably involves both social values and economic possibilities. In practice, however, their estimates of the amount of industrial employment required in new towns or substantial extensions of existing towns are too often based on the application of national percentages to the local population. The use of indices of industrial diversity for this particular purpose would produce the same result. Since the form of index which is most generally used is based on the deviations of a town's industrial structure from the national average, attempts to maximise the index for any town are equivalent to attempts to reproduce the national economy within it. Nevertheless it would seem unreasonable to suppose that it is either practicable or desirable to do so.

Indices of social mobility are intended to measure the extent to which sons choose freely from the occupations open to them. An index of unity indicates a state of affairs in which there is no association between the occupations chosen by the sons and those in which their fathers are engaged. The authors of this index state specifically that they consider such circumstances to be neither ideal nor practicable; nevertheless such statements have not prevented some educationalists from concluding that one of the aims of education should be to maximise the index.

The determination of statistical standards which can be used as guides to policy in either the town planning or the educational fields obviously presents many difficulties.

The administrator's problems are of two kinds: he needs to determine both the aims of his policy and the means to achieve them. Though he may need statistics which are standards for the solution of the first group, he will certainly need statistics which are measurements for the technical problems which form the second group. Also since the phenomena being considered are complex, satisfactory measurements for their description will be possible only within a framework of social theory. It is not unreasonable to argue that indices of industrial diversity and social mobility cannot be used to provide satisfactory standards, since they are not satisfactory measurements. Both measure deviations from impossible situations! On the other hand, the development of social theory in these particular instances may not be possible without a simultaneous development in social statistics.

Perhaps the social statistician's most useful contribution to the

solution of these problems would be the construction of a variety of indices to measure various aspects of these phenomena. Their plurality might well prevent their being used as false standards, and at the same time their existence might assist the formulation of a more comprehensive social theory. In constructing these indices, the social statistician should be primarily concerned with social theory and not with social values, since his concern should be to produce satisfactory measurements; but in this case, as in the case of the simpler problems discussed in the first part of this paper, consideration of the interests of administrators may well help to determine those aspects of the situation which are significant and should be measured.

BIBLIOGRAPHY

Cole, G. D. H. (1957). 'Sociology and social policy', *British Journal of Sociology*, **8**, 158.

Booth, C. (1887). 'The inhabitants of the Tower Hamlets', *J. R. Statist. Soc.*, **50**, 326.

Rowntree, B. S. (1902). *Poverty, A Study of Town Life.* Macmillan, London.

Bowley, A. L. and Hurst, A. R. B. (1915). *Livelihood and Poverty.* Bell, London.

Caradog-Jones, D. (Editor). (1934). *A Social Survey of Merseyside.* Liverpool University Press.

Townsend, P. (1954). 'Measuring poverty', *British Journal of Sociology*, **5**, 130.

International Definition and Measurement of Standards and Levels of Living (1954). UN, New York.

Census of England and Wales, 1931. Housing, Report and Tables. (1935). HMSO, London.

Census 1951. England and Wales. Housing Report. (1956). HMSO, London.

Simon, E. D. (1929). *How to Abolish Slums.* Longmans Green, London.

Tress, R. C. (1938). 'Unemployment and diversification of industry', *Manchester School*, **9**, 140.

Glass, D. V. (Editor). (1954). *Social Mobility in Britain.* Routledge, London.

Conway, F. (1953). 'The industrial structure of towns', *Manchester School*, **21**, 154.

Dunning, J. H. (1960). 'Manufacturing industry in the New Towns', *Manchester School*, **28**, 137.

Billewicz, W. Z. (1955). 'Some remarks on the measurement of social mobility', *Population Studies*, **9**, 96.

Musgrove, F. (1959). 'Background and development', *Times Educational Supplement*, September 18, 278.

SOME NOTES ON THE PRECISION OF THE GRADIENT OF AN ESTIMATED STRAIGHT LINE

DONALD G. BEECH
British Ceramic Research Association, Stoke-on-Trent

In this article Dr Beech assembles some material concerning the confidence limits to be attached to the estimated gradients of linear functional relationships and compares them with those for the regression lines. He derives explicit formulae for these confidence limits, and also discusses the accuracy to which the relative error of x and y needs to be known when the gradient is being estimated. The more important formulae are summarised at the end and mathematical details are given in an appendix.

Introduction

It is quite a simple matter to calculate the standard error of the gradient of a regression line and hence to give its confidence limits; formulae for this are well known. The calculation of the gradient of the functional relationship between x and y when the estimated values of x and y are both subject to error is also fairly simple if the ratio of these experimental errors is known and is constant. However, the calculation of confidence limits for this estimate is rather more troublesome and the appropriate methods are not so well known. It is the purpose of the present article to bring together some of the relevant material, to show how a certain amount of unification is possible, and to present some approximate formulae. In addition, some evidence is given concerning the magnitude of the changes in the calculated gradient produced by changes in the assumed ratio of experimental errors.

Regression Lines

If we consider our basic variables to be x and y, then for simplicity we may define

$$s_{xx} = \Sigma\,(x-\bar{x})^2 \qquad s_{yy} = \Sigma\,(y-\bar{y})^2 \qquad s_{xy} = \Sigma\,(x-\bar{x})(y-\bar{y})$$

The gradient b_1 of the regression of y on x is given by

$$b_1 = \frac{s_{xy}}{s_{xx}}$$

and the residual variance s^2 is given by $\left(s_{yy} - \frac{s_{xy}^2}{s_{xx}}\right) \div (n-2)$, where n is the number of pairs of observations. The variance of b_1 is usually given in the form s^2/s_{xx} but for our present purpose it is convenient to consider the ratio $\mathrm{V}(b_1)/b_1^2$. This is seen to be

$$s^2 \frac{s_{xx}}{s_{xy}^2}$$

$$= \left(s_{yy} - \frac{s_{xy}^2}{s_{xx}} \right) \frac{s_{xx}}{(n-2)s_{xy}^2}$$

$$= \left(\frac{s_{xx}s_{yy}}{s_{xy}^2} - 1 \right) \div (n-2)$$

$$= \frac{1-r^2}{r^2(n-2)}$$

where r is the correlation coefficient between x and y. It is clear that for given values of r and n the fractional standard error of b_1 is determined. In fact

$$\frac{\text{S.E.}(b_1)}{b_1} = \sqrt{\left[\frac{1-r^2}{r^2(n-2)} \right]}$$

Let this quantity be termed K. The only advantages in this way of regarding the problem are (1) that it enables a rough estimate of the number of observations necessary for a given fractional standard error to be made, provided that the degree of correlation is roughly known; (2) that it forms a link unifying the several sections of this article. Confidence limits for b_1 are given by $b_1(1+Kt)$, where Student's t may take its positive or negative values for $(n-2)$ degrees of freedom and the required level of confidence.

If we consider b_2, the gradient $\mathrm{d}x/\mathrm{d}y$ of the regression of x on y, then $b_2 = s_{xy}/s_{yy}$;

$$\mathrm{V}(b_2) = \frac{(s')^2}{s_{yy}} \quad \text{where} \quad (s')^2 = \left(s_{xx} - \frac{s_{xy}^2}{s_{yy}} \right) \div (n-2)$$

$$\frac{\mathrm{V}(b_2)}{{b_2}^2} = \frac{s_{yy}}{s_{xy}^2} \left(s_{xx} - \frac{s_{xy}^2}{s_{yy}} \right) \div (n-2)$$

$$= \frac{1-r^2}{r^2(n-2)}$$

$$\therefore \quad \frac{\text{S.E.}(b_2)}{b_2} = K$$

The identity in form of this equation and the one for the other regression is interesting; it is not unexpected, of course, since the correlation coefficient, which determines K for a given n, does not depend on which variable is considered as the independent one. The confidence limits for b_2 are given by $b_2(1+Kt)$.

Because the quantities K and Kt are of some importance, nomograms have been prepared from which they can be estimated for values of r in the range 0·7-0·999 and of n in the range 6-100. Fig. 1 gives the value of K, Fig. 2 gives Kt for a 95% confidence interval, and Fig. 3

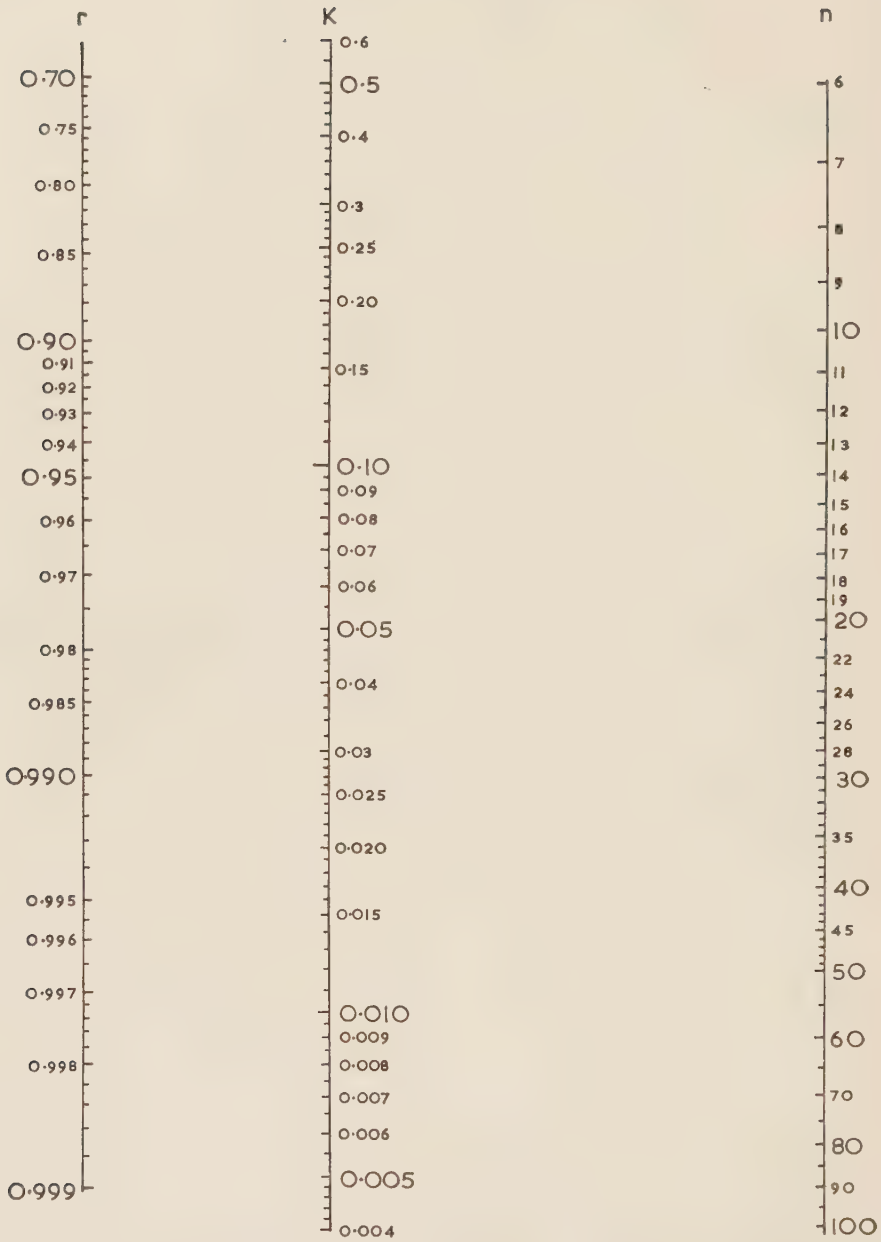


FIG. 1. Nomogram for quantity K defined in the text. A straight-edge joining r, the correlation coefficient, and n, the number of pairs of observations, cuts the K-scale at the required value.

gives Kt for a 99% confidence interval. The accuracy of the nomograms is sufficient for normal purposes.

The identity of the expressions for the fractional standard error of the two regressions leads one to wonder whether some similar relationship may hold for the linear functional relationship (if it exists), the

FIG. 2. Nomogram for Kt from which 95% confidence intervals for gradients may be found. A straight-edge joining r, the correlation coefficient, and n, the number of pairs of observations, cuts the Kt-scale at the required value.

gradient of which will be between those of the two regressions whatever the relative magnitudes of the x and y errors.

Functional Relationships: General Formulae

Consider the variables x, y, which have normally distributed and uncorrelated errors. The error variances, $V(y)$ and $V(x)$, are constant and the ratio $V(y)/V(x)$ has the known value λ. It is required to estimate the gradient of the graph of y against x; let this gradient be called b.

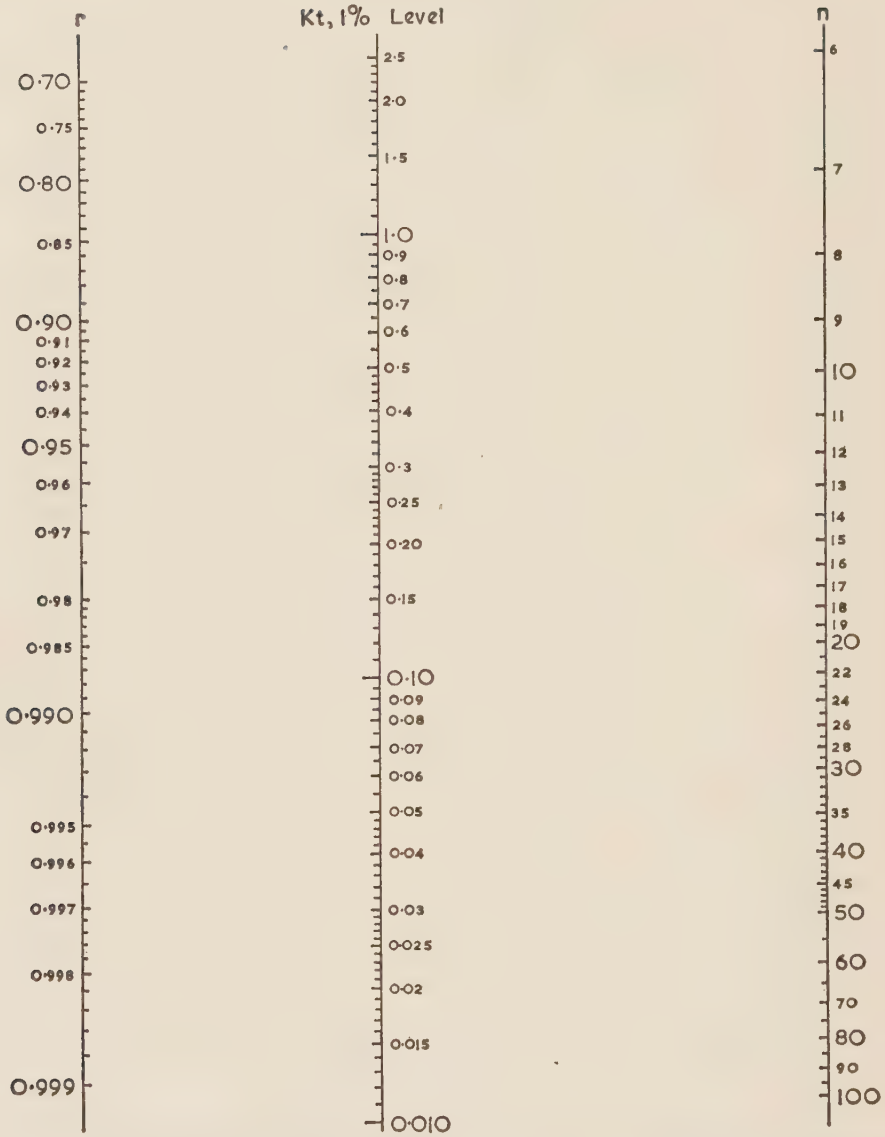


FIG. 3. Nomogram for Kt from which 99% confidence intervals for gradients may be found. A straight-edge joining r, the correlation coefficient, and n, the number of pairs of observations, cuts the Kt-scale at the required value.

It is more convenient in the derivation of formulae to work with variables z and y, z being defined as $x\sqrt{(\lambda)}$.

Then
$$\frac{V(y)}{V(z)} = \frac{V(y)}{\lambda V(x)} = 1$$

$$s_{zz} = s_{xx}\lambda \qquad s_{zy} = s_{xy}\sqrt{(\lambda)}$$

The gradient of the y-z graph, B, is related to b by the equation $b = B\sqrt{(\lambda)}$. It is convenient to take the positive or negative root of λ

according as s_{xy} is positive or negative; then s_{zy} and B are always positive.

E. J. Williams (reference 1, p. 198) has tackled the problem of estimating the gradient by working in terms of a linear function

$$\xi = y + \gamma z$$

which is a quantity affected only by random errors. It is easy to see that the quantity γ, which he then estimates, is the negative of B. (I follow Williams here because he goes on to deduce fiducial limits, but I have altered some of his symbols so as to make the symbolism in this article more self-consistent.)

It may be shown that the best estimate of γ is given by

$$\gamma = -\{w + \sqrt{(w^2+1)}\}$$

where

$$w = \frac{s_{yy} - s_{zz}}{2s_{zy}}$$

This reduces to the equation

$$b = \theta \pm \sqrt{(\theta^2 + \lambda)}$$

where $\theta = \dfrac{s_{yy} - \lambda s_{xx}}{2s_{xy}}$ and the sign of the root is the same as the sign of s_{xy}. This result has been deduced by Lindley[2] and others.

Williams goes on to show that fiducial limits for γ may be obtained by solving

$$[\gamma^2 s_{zy} + \gamma(s_{yy} - s_{zz}) - s_{zy}]^2 = \frac{F(1+\gamma^2)^2}{n-2}(s_{zz}s_{yy} - s^2{}_{zy})$$

where F is the value of the variance ratio at the required significance level for $(n-2)$ degrees of freedom. This equation is not very easy to work with, because it has four roots, two of which are the ones required and the other two apply to the worst estimate of the gradient rather than to the best.

However, it is possible [as shown in Appendix (*a*)] to reduce this equation to a form which is somewhat simpler to use. The fiducial limits are given by

$$\mathrm{L}(\gamma) = -\{w + \sqrt{(w^2 + 1 - K^2t^2)}\} \div (1 - Kt)$$

where K is defined as before. (The change of variable leaves K unaffected.) The two correct limits are obtained by giving Student's t its appropriate positive and negative values. The value of B is the same as that of γ but the sign is changed to positive.

A different method of determining confidence limits for B was published in 1956 by Miss Monica Creasy.[3] It is interesting to note that the method, although superficially dissimilar, gives results identical with those obtained by Williams's method. A demonstration of this identity is given in Appendix (*b*). The method is not quoted here because I do not find it convenient to use.

It is simple to transform the above equation into one giving limits for b. (Williams calls his limits 'fiducial limits', and Miss Creasy calls hers 'confidence limits'; hereafter I shall usually write of them as 'limits' or 'confidence limits'.) We have

$$L(b) = \{\theta \pm \sqrt{[\theta^2 + \lambda(1 - K^2t^2)]}\} \div (1 - Kt)$$

where the sign of the square root is taken as the sign of s_{xy}. This equation is convenient to use, because the quantities θ and λ are used in determining b, and K is quite simply obtained, either by calculation or from the nomogram in Fig. 1. Again, t is the value for the required significance level with $(n-2)$ degrees of freedom, and is counted as positive or negative to furnish the two limits. A range of values of Kt for 95% and 99% limits may be obtained from Figs. 2 and 3.

It is clear that, unlike the limits for regression lines, the limits here will be asymmetric with respect to the best estimate, b. As K decreases, however, the asymmetry becomes less marked, and for reasonably precise estimates it is not very pronounced.

For comparison with the case of the regression lines we may calculate the quantity: half the width of the confidence interval divided by the midpoint of the interval. Let the limit when t is given its positive value be L_1 and the other value L_2. Consider only the case when s_{xy} is positive; similar reasoning applies when it is negative.

Let $$\theta + \sqrt{[\theta^2 + \lambda(1 - K^2t^2)]} = M$$

$$\frac{L_1 - L_2}{2} = \frac{M}{2}\left(\frac{1}{1 - K|t|} - \frac{1}{1 + K|t|}\right) = \frac{MK|t|}{1 - K^2t^2}$$

$$\frac{L_1 + L_2}{2} = \frac{M}{2}\left(\frac{1}{1 - K|t|} + \frac{1}{1 + K|t|}\right) = \frac{M}{1 - K^2t^2}$$

Hence $$\left(\frac{L_1 - L_2}{2}\right) \div \left(\frac{L_1 + L_2}{2}\right) = K|t|$$

This is precisely the result that would be obtained by applying the same process to the regression lines. It is clear that if the slight asymmetry is ignored, the relative precision of the functional relationship is the same as that of the regression lines.

To conclude this section it may be pointed out that the equation for the limits is satisfactory provided that Kt is not too large. If Kt increases, the stage is eventually reached when the confidence regions of the best straight line and the worst straight line overlap. (The worst straight line is the line of minimum likelihood.) I shall not dwell on that situation here, because if a functional relationship is to be sufficiently accurate to be useful it is necessary for Kt to be small compared with 1, say $\gtrsim 0{\cdot}1$.

Functional Relationships: Approximate Formulae

It is possible to derive approximate formulae for the limits which are

quite accurate enough when Kt is small and which are somewhat easier to use than the general formulae. From the equation in the previous section concerning the width of the confidence interval it is clear that, if the asymmetry is only slight, the fractional precision of the estimate b is of the order of Kt. (This can be shown more rigorously but it is sufficient to state it here.) As already stated, whatever the level of confidence we employ, our estimate b is not worth much unless Kt is a small fraction. I shall therefore assume here the situation where $Kt \leqslant 0{\cdot}1$. In Appendix (*c*) it is shown that in these circumstances

$$\frac{K^2t^2}{w^2+1} \leqslant 0{\cdot}01$$

and it is possible by some simple algebra to show that a good approximate formula is

$$\frac{\mathrm{L}(B)}{B} = \frac{1-[K^2t^2/(B^2+1)]}{1-Kt}$$

Since

$$\frac{\mathrm{L}(B)}{B} = \frac{\mathrm{L}(b)}{b} \quad \text{and} \quad b = B\sqrt{(\lambda)}$$

$$\frac{\mathrm{L}(b)}{b} = \frac{1-\left[K^2t^2\Big/\left(\frac{b^2}{\lambda}+1\right)\right]}{1-Kt}$$

where t is given its positive and negative values.

The error in this estimate of the limit is appreciably less than 0·01% when $Kt = 0{\cdot}1$, which is negligible considering that the limit itself will then be about 10% distant from b. For smaller values of Kt the approximation is even better.

In a similar manner it may be shown that the distance of the limit from b is given by the formula

$$\frac{\mathrm{L}(b)-b}{b} = Kt\left\{\frac{1-\left[Kt\Big/\left(\frac{b^2}{\lambda}+1\right)\right]}{1-Kt}\right\}$$

This supports the earlier statement that the fractional precision of the estimate b is of the order of Kt.

Functional Relationship through Origin

If we seek to fit a line $y = b'x$ to a set of points subject to error in both x and y, a somewhat similar procedure may be followed. (This method is stated by Miss Creasy[3] to be appropriate provided that the values x, y are regarded as estimates of fixed true values.) The gradient b' is given by

$$b' = \theta' \pm \sqrt{(\theta'^2+\lambda)}$$

where

$$\theta' = \frac{\Sigma y^2 - \lambda \Sigma x^2}{2\,\Sigma xy}$$

Exactly similar formulae for the limits are used with the substitution of θ' for θ and K' for K. K' is now defined as

$$\sqrt{\left[\frac{\Sigma x^2 \Sigma y^2}{(\Sigma xy)^2} - 1\right]}$$

and t has $(n-1)$ degrees of freedom.

Sensitivity of Functional Relationship to Relative Errors of x and y

The formula for the gradient of the linear functional relationship is, as stated in a previous section,

$$b = \frac{s_{yy} - \lambda s_{xx}}{2s_{xy}} + \sqrt{\left[\left(\frac{s_{yy} - \lambda s_{xx}}{2s_{xy}}\right)^2 + \lambda\right]}$$

when s_{xy} is positive and $\lambda = \mathrm{V}(y)/\mathrm{V}(x)$, i.e. the ratio of the squares of the experimental errors of y and x. The question at issue is: how sensitive is the value of b to λ? Since the value of λ will rarely be known accurately, this is a point of some importance. It may be answered fairly readily in specific cases.

The general question is difficult to answer directly, owing to the fact that there are four variables in the equation. However, it would seem reasonable to consider as important the ratio of the experimental error to the standard deviation of the observed values. This proves to be the case. [For details see Appendix (*d*).]

If we define $$C^2 = \frac{\mathrm{V}(y)/s_{yy}}{\mathrm{V}(x)/s_{xx}} \quad \text{and} \quad g^2 = \frac{s_{xx}}{s_{yy}}$$

$$b = \frac{1 - C^2}{2gr} + \sqrt{\left[\left(\frac{1 - C^2}{2gr}\right)^2 + \lambda\right]}$$

where r is the correlation coefficient. If $\bar{b}$ is the value when $C = 1$, then $\bar{b} = 1/g$ and

$$\frac{b}{\bar{b}} = \frac{1 - C^2}{2r} + \sqrt{\left[\left(\frac{1 - C^2}{2r}\right)^2 + C^2\right]}$$

This measure of the proportional variation of b is easy to manipulate and can be used to determine the effect of varying C from 0 to ∞ for fixed values of r. What has actually been done in the following Tables is to take four typical values for n, the number of points on the graph, ranging from $n = 10$, a small number of points, to $n = 100$, a fairly large number, and for each value to calculate the values of r which are necessary to make $Kt_{0\cdot05}$ equal to 0·1, 0·05, and 0·01. For each set of values of r, the values of $b/\bar{b}$ for various values of C^2 were then calculated from the above equation. The results are given in the form $(b - \bar{b})/\bar{b}$ in Tables I (*a-c*). The reason for this presentation is that if the values of the 95% confidence limits for b are denoted by $\mathrm{L}_1(b)$, $\mathrm{L}_2(b)$ then $|\mathrm{L}_1(b) - b|/b$ and $|\mathrm{L}_2(b) - b|/b$ are each of the order of $Kt_{0\cdot05}$. For the

fairly narrow range of values of $b/\bar{b}$ covered by the Tables it is also true that $|L_1(b) - b|/\bar{b}$ and $|L_2(b) - b|/\bar{b}$ are practically of the order of $Kt_{0\cdot05}$.

TABLE I

Sensitivity of b to variation in C^2

(*a*) $Kt_{0\cdot05} \simeq 0\cdot1$

n	r	$(b-\bar{b})/\bar{b}$ for Following Values of C^2:							
		0	0·1	0·2	0·5	2	5	10	∞
10	0·992	$\times 10^{-4}$ 81	$\times 10^{-4}$ 66	$\times 10^{-4}$ 54	$\times 10^{-4}$ 27	$\times 10^{-4}$ −27	$\times 10^{-4}$ −53	$\times 10^{-4}$ −66	$\times 10^{-4}$ −80
25	0·975	256	210	171	86	−85	−169	−206	−250
50	0·940	638	525	429	214	−210	−421	−508	−610
100	0·900	1111	916	749	375	−362	−697	−839	−1000

(*b*) $Kt_{0\cdot05} \simeq 0\cdot05$

n	r	$(b-\bar{b})/\bar{b}$ for Following Values of C^2:							
		0	0·1	0·2	0·5	2	5	10	∞
10	0·998	$\times 10^{-4}$ 20	$\times 10^{-4}$ 16	$\times 10^{-4}$ 13	$\times 10^{-4}$ 7	$\times 10^{-4}$ −7	$\times 10^{-4}$ −13	$\times 10^{-4}$ −16	$\times 10^{-4}$ −20
25	0·993	70	58	47	23	−23	−47	−57	−70
50	0·985	152	125	102	51	−50	−101	−123	−150
100	0·970	309	254	207	103	−103	−203	−248	−300

(*c*) $Kt_{0\cdot05} \simeq 0\cdot01$

n	r	$(b-\bar{b})/\bar{b}$ for Following Values of C^2:							
		0	0·1	0·2	0·5	2	5	10	∞
10	0·9999	$\times 10^{-5}$ 10	$\times 10^{-5}$ 8	$\times 10^{-5}$ 7	$\times 10^{-5}$ 3	$\times 10^{-5}$ −3	$\times 10^{-5}$ −7	$\times 10^{-5}$ −8	$\times 10^{-5}$ −10
25	0·9997	30	25	20	10	−10	−20	−25	−30
50	0·9994	60	49	40	20	−20	−40	−49	−60
100	0·9988	120	98	80	40	−40	−80	−98	−120

Hence, for each Table, the ratio of the value of $(b - \bar{b})/\bar{b}$ to the value of $Kt_{0\cdot05}$ is roughly equal to the ratio of $(b - \bar{b})$ to $|L(b) - b|$, i.e. the ratio of $(b - \bar{b})$ to the precision of estimation of b. It seems reasonable to suggest that unless $(b - \bar{b})$ is appreciably more than one-tenth of the precision of the estimate of b, the difference between b and $\bar{b}$ is not very important. On this basis, the differences above the bold lines in Tables I (*a-c*) may be regarded as of little importance. It is clear that in some circumstances C can cover quite a wide range without producing important changes in b. When the 'region of no importance', for a

given number of points and value of Kt, covers the whole range of values of C from 0 to ∞, then, if we have insufficient information even to make a guess at C, it would be quite justifiable to take the value of $\bar{b}$ as a good estimate of b. (Examples of this situation are $n=10$, $r=0{\cdot}992$, $Kt\simeq 0{\cdot}1$; $n=25$, $r=0{\cdot}9997$, $Kt\simeq 0{\cdot}01$.) The estimate $\bar{b}$ is very simply obtained, as $\bar{b}=1/g=\sqrt{(s_{yy}/s_{xx})}$. This estimate of the gradient of the functional relationship has at times been proposed; it is interesting to note these conditions in which it may be practically justifiable. Some of the correlation coefficients quoted in the Tables may seem high to the statistician. However, I think that values as high as these will often be found if the results of physical measurements are analysed. In my own work I have encountered values higher than any in the Tables.

It is also interesting to note that, for a given number of points, the better the correlation the less necessary it is to strain after very precise estimates of C (and therefore of λ). To some extent this surprised me, but it is an additional argument in favour of extra care in making experimental measurements. Conversely, it would be more necessary to know the value of C precisely at higher values of Kt; the Tables have not been extended to cover these values because I consider that, if the sampling error of the gradient estimate is greater than 10%, the estimate is of little use for most functional relationships.

The foregoing should not be taken to imply that no steps should be taken to determine the appropriate value of C. On the contrary, the

TABLE II

Sensitivity of b to variation in C^2:

(*a*) $Kt_{0{\cdot}05} \simeq 0{\cdot}1$

n	r	(Difference in Values of $b)/\bar{b}$ for Following Values of C^2:							
		0-0·1	0·1-0·2	0·2-0·5	0·5-1	1-2	2-5	5-10	10-∞
		$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$
10	0·992	15	12	27	27	27	26	13	14
25	0·975	46	39	85	86	85	84	37	44
50	0·940	113	96	215*	214*	210*	211*	87	102
100	0·900	195*	167*	374*	375*	362*	335*	142*	161*

(*b*) $Kt_{0{\cdot}05} \simeq 0{\cdot}05$

n	r	(Difference in Values of $b)/\bar{b}$ for Following Values of C^2:							
		0-0·1	0·1-0·2	0·2-0·5	0·5-1	1-2	2-5	5-10	10-∞
		$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$
10	0·998	4	3	6	7	7	6	3	4
25	0·993	12	11	24	23	23	24	10	13
50	0·985	27	23	51	51	50	51	22	27
100	0·970	55	47	104*	103*	103*	100*	45	52

(*c*) $Kt_{0\cdot05} \simeq 0\cdot01$

n	r	(Difference in Values of $b/\bar{b}$ for Following Values of C^2:							
		0-0·1	0·1-0·2	0·2-0·5	0·5-1	1-2	2-5	5-10	10-∞
10	0·9999	$\times 10^{-5}$ 2	$\times 10^{-5}$ 1	$\times 10^{-5}$ 4	$\times 10^{-5}$ 3	$\times 10^{-5}$ 3	$\times 10^{-5}$ 4	$\times 10^{-5}$ 1	$\times 10^{-5}$ 2
25	0·9997	5	5	10	10	10	10	5	5
50	0·9994	11	9	20	20	20	20	9	11
100	0·9988	22	18	40	40	40	40	18	22

* Indicates differences of practical importance.

best possible estimate of C should be made in the light of the information available. But when this has been said, it must be admitted that the value of C will rarely be known precisely. More often we can only give the likely range of values and sometimes can make only a very rough guess. Tables II (*a-c*) are given in the hope that the guesses may be made with some knowledge of the consequences. They have been calculated from Table I by taking the differences between the values of $(b-\bar{b})/\bar{b}$ given by adjacent values of C^2 at the head of each Table; they therefore give the difference between adjacent values of b in terms of $\bar{b}$. Each such difference may be considered important if it appreciably exceeds the appropriate value of $0\cdot1Kt_{0\cdot05}$, for the reason given in connection with the values of $(b-\bar{b})/\bar{b}$ in Table I. The important differences have been marked with an asterisk. Hence in the parts of the Tables free from asterisks a greater latitude in the value of C is allowable.

Thus if $n=25$, $r=0\cdot993$, $Kt\simeq0\cdot05$, and we think that C^2 is roughly 0·2, we can proceed with some confidence because Table II*b* shows that C^2 can vary from 0 to 1 without seriously affecting the estimate of b. (Here we add adjacent cells: $0\cdot0012+0\cdot0011=0\cdot0023$, and $0\cdot0024+0\cdot0023=0\cdot0047$.) Also if $n=50$, $r=0\cdot9994$, $Kt\simeq0\cdot01$, suppose that we are reasonably confident that C^2 is greater than 10, but do not know its exact value. Table II*c* shows that it is quite adequate to assume $C^2=\infty$, i.e. to use the regression of y upon x. The use of one or other of the regression lines as an estimate of the functional relationship is often defended on the grounds that one of the variables is much more precisely measured than the other. Table II may help in giving a quantitative meaning to 'much more precisely'.

Similar calculations may be carried out for a functional relationship that is forced to pass through the origin. The formula for the gradient has been given in an earlier section (p. 21). The same formula that is used for $b/\bar{b}$ can be employed here, the value of C^2 being taken as

$$\frac{V(y)/\Sigma y^2}{V(x)/\Sigma x^2}$$

and the value of r being replaced by $\Sigma xy/\sqrt{(\Sigma x^2\Sigma y^2)}$. The quantity

$K't$ has to be calculated; K' is as defined earlier, and t has $(n-1)$ degrees of freedom.

Tables I and II are applicable to this situation if they are slightly modified as follows. The numbers of points to which the values in the body of the Tables apply are 9, 24, 49, and 99 instead of 10, 25, 50, and 100; r is replaced by $\Sigma xy/\sqrt{(\Sigma x^2 \Sigma y^2)}$.

Summary of More Important Working Rules

In the following equations the variables are x and y and there are n pairs of values. The precision of estimation of x and y is constant and the relative precision is indicated by the quantity λ, where

$$\lambda = (\text{error variance of } y) \div (\text{error variance of } x)$$

Other convenient symbols are:

$$s_{xx} = \Sigma(x-\bar{x})^2$$
$$s_{yy} = \Sigma(y-\bar{y})^2$$
$$s_{xy} = \Sigma(x-\bar{x})(y-\bar{y})$$
$$r = \text{correlation coefficient} = s_{xy}/\sqrt{(s_{xx}s_{yy})}$$
$$K = \sqrt{\left[\frac{1-r^2}{r^2(n-2)}\right]}$$
$$t = \text{Student's } t, \text{ obtainable from Tables.}$$

For certain ranges of n and r, K, $Kt_{0\cdot05}$, and $Kt_{0\cdot01}$ may be approximately determined from the nomograms (Figs. 1-3).

Gradient of Regression of y on x (b_1)

$$b_1 = s_{xy} \div s_{xx}$$

Confidence limits are given by

$$b_1(1+Kt)$$

where t is the value with $(n-2)$ degrees of freedom at the required level of confidence; t may be either positive or negative, thus providing the two limits.

Gradient of Functional Relationship (b)

$$b = \theta + \sqrt{(\theta^2+\lambda)} \qquad \text{when } s_{xy} \text{ is positive}$$
$$b = \theta - \sqrt{(\theta^2+\lambda)} \qquad \text{when } s_{xy} \text{ is negative}$$

where

$$\theta = \frac{s_{yy} - \lambda s_{xx}}{2s_{xy}}$$

Confidence limits, $\mathrm{L}(b)$, are given by

$$\mathrm{L}(b) = \{\theta + \sqrt{[\theta^2 + \lambda(1-K^2t^2)]}\} \div (1-Kt)$$
when s_{xy} is positive

$$\mathrm{L}(b) = \{\theta - \sqrt{[\theta^2 + \lambda(1-K^2t^2)]}\} \div (1-Kt)$$
when s_{xy} is negative.

Again t has $(n-2)$ degrees of freedom and the two limits are obtained by giving t its positive and negative values. Approximate values of $L(b)$ are given by

$$\frac{L(b)}{b} = \frac{1-\left[K^2t^2\Big/\left(\frac{b^2}{\lambda}+1\right)\right]}{1-Kt}$$

Again t has $(n-2)$ degrees of freedom and takes positive and negative values. The approximation is satisfactory provided that $Kt \leqslant 0{\cdot}1$.

Functional Relationship Constrained through Origin

The equations for the gradient (b') and for its confidence limits have the same forms as those applicable to the general functional relationship, but with the following changes in symbols:

θ is replaced by θ', where $\theta' = \dfrac{\Sigma y^2 - \lambda \Sigma x^2}{2\,\Sigma xy}$

K is replaced by K', where $K' = \sqrt{\left[\dfrac{\Sigma x^2 \Sigma y^2}{(\Sigma xy)^2} - 1\right]}$

t has $(n-1)$ degrees of freedom

b is replaced by b' in the approximate equation for the confidence limits.

Acknowledgement

I wish to express my thanks to Dr N. F. Astbury, Director of Research of the British Ceramic Research Association, for permission to publish this paper.

APPENDIX

(*a*) *Simplification of Formula for Limits*

Williams[1] has shown that fiducial limits for γ are obtained by solving

$$[\gamma^2 s_{zv} + \gamma(s_{vv} - s_{zz}) - s_{zv}]^2 = \frac{F(1+\gamma^2)^2(s_{zz}s_{vv} - s_{zv}^2)}{n-2}$$

Hence $$\left[\frac{\gamma^2}{2} + \gamma\left(\frac{s_{vv}-s_{zz}}{2s_{zv}}\right) - \tfrac{1}{2}\right]^2 = \frac{F}{n-2}\left(\frac{1+\gamma^2}{2}\right)^2\left(\frac{s_{zz}s_{vv}-s_{zv}^2}{s_{zv}^2}\right)$$

Put $$\frac{s_{vv}-s_{zz}}{2s_{zv}} = w$$

$$\left(\frac{\gamma^2}{2} + \gamma w - \tfrac{1}{2}\right)^2 = F\left(\frac{1+\gamma^2}{2}\right)^2 \cdot \frac{1-r^2}{(n-2)r^2}$$

$$\frac{\gamma^2}{2}+\gamma w-\tfrac{1}{2} = \sqrt{(F)}\cdot\left(\frac{1+\gamma^2}{2}\right)K$$

$$= Kt\left(\frac{1+\gamma^2}{2}\right)$$

where K is defined as in the main text. (The value of r is the same for z and y as it is for the untransformed variables x and y, and so, therefore, is K.) Any choice of sign on taking the square root is accommodated later on by allowing t to have its positive or negative value.

$$\frac{\gamma^2}{2}(1-Kt)+\gamma w-\tfrac{1}{2}(1+Kt) = 0$$

If we call the relevant solution $\mathrm{L}(\gamma)$, then

$$\mathrm{L}(\gamma) = -\{w^2+\sqrt{(w^2+1-K^2t^2)}\}\div(1-Kt)$$

The positive square root gives the appropriate root of the quadratic equation, as is clearly seen by putting $t=0$, when the equation for the best estimate of γ is obtained (for the case where s_{zy} is positive, as it is here).

(*b*) *Identity of Two Methods for Confidence Limits*

Miss Creasy has given an expression for the calculation of confidence limits in terms of a transformed variable ϕ. She puts $\phi=\arctan$ (gradient) and takes $\lambda=1$. Then the appropriate equation, transferred to our symbols and with a simple algebraic rearrangement, is:

$$\sin 2(\hat{\phi}-\phi) = \frac{Kt}{\sqrt{(w^2+1)}}$$

where $\hat{\phi}$ is the maximum-likelihood value and ϕ gives the limits by substituting positive and negative values of t. The value of $\tan\hat{\phi}$ is readily found directly, or by Lindley's or Williams's equations, to be

$$w+\sqrt{(w^2+1)}$$

for the case where s_{zy} is positive. There is no trouble in isolating the correct limits provided that $(\hat{\phi}-\phi)<45°$, which I shall assume to be the case. Hence it may be shown that

$$\tan 2\hat{\phi} = -\frac{1}{w}$$

$$w\sin 2\hat{\phi} = -\cos 2\hat{\phi} \qquad \ldots\ (1)$$

(This may also be shown to be true if s_{zy} is negative.)

From Williams's approach the equation

$$\frac{\gamma^2}{2}+\gamma w-\tfrac{1}{2} = Kt\left(\frac{1+\gamma^2}{2}\right)$$

deduced in Appendix (*a*) may be taken and transformed, by putting $\gamma = -\tan\phi$, to

$$\tfrac{1}{2}\tan^2\phi - w\tan\phi - \tfrac{1}{2} = \frac{Kt}{2}\sec^2\phi$$

$$\tfrac{1}{2}\sin^2\phi - w\sin\phi\,.\cos\phi - \tfrac{1}{2}\cos^2\phi = \frac{Kt}{2}$$

$$w\sin 2\phi + \cos 2\phi = -Kt \qquad \ldots\ldots (2)$$

Now $\sin 2(\hat{\phi} - \phi) = \sin 2\hat{\phi}\,.\cos 2\phi - \cos 2\hat{\phi}\,.\sin 2\phi$

From equation (1)

$$\sin 2(\hat{\phi} - \phi) = \sin 2\hat{\phi}\,.\cos 2\phi + w\sin 2\hat{\phi}\,.\sin 2\phi$$

$$= \sin 2\hat{\phi}\,(\cos 2\phi + w\sin 2\phi)$$

From equation (2)

$$\sin 2(\hat{\phi} - \phi) = -Kt\sin 2\hat{\phi}$$

But, by squaring equation (1)

$$w^2\sin^2 2\hat{\phi} = \cos^2 2\hat{\phi} = 1 - \sin^2 2\hat{\phi}$$

$$\therefore \quad \sin 2\hat{\phi} = \frac{1}{\sqrt{(w^2+1)}}$$

Hence $$\sin 2(\hat{\phi} - \phi) = -\frac{Kt}{\sqrt{(w^2+1)}}$$

This expression for the limits, derived from Williams's equation, is identical with that from Miss Creasy's apart from a change of sign, which is immaterial since t has to be given its positive and negative values.

(*c*) *Approximate Formulae for Limits*

It is most convenient to work in terms of the gradient of the graph of y against z. Then this gradient B is given by

$$B = w + \sqrt{(w^2+1)}$$

If we put $$v^2 = w^2 + 1$$

then $$B = w + v$$

$$\mathrm{L}(B) = \frac{w + \sqrt{(v^2 - K^2t^2)}}{1 - Kt}$$

$$= \frac{w + v\left(1 - \dfrac{K^2t^2}{v^2}\right)^{\frac{1}{2}}}{1 - Kt}$$

If we confine our attention to the case where $Kt \leqslant 0{\cdot}1$ then, since v is not less than 1, $K^2t^2/v^2 \leqslant 0{\cdot}01$. We can therefore expand the square root and obtain

$$\mathrm{L}(B) = \frac{w + v\left(1 - \frac{K^2t^2}{2v^2} - \frac{K^4t^4}{8v^4} \cdots\right)}{1 - Kt}$$

Hence

$$\frac{\mathrm{L}(B)}{B} = \frac{1 - \frac{K^2t^2}{2v(w+v)} - \frac{K^4t^4}{8v^3(w+v)} \cdots}{1 - Kt}$$

Now

$$\begin{aligned} B^2 &= w^2 + 2wv + v^2 \\ B^2 + 1 &= w^2 + 1 + 2wv + v^2 \\ &= 2v(w+v) \\ (B^2+1)^3 &= 8v^3(w+v)^3 \\ \frac{(B^2+1)^3}{B^2} &= 8v^3(w+v) \end{aligned}$$

Therefore

$$\frac{\mathrm{L}(B)}{B} = \frac{1 - \frac{K^2t^2}{B^2+1} - \frac{B^2K^4t^4}{(B^2+1)^3} \cdots}{1 - Kt}$$

It is clear that the third term in the numerator cannot be greater than K^4t^4 $(=0{\cdot}0001)$; actually it is appreciably less than 0·0001. Now if $Kt = 0{\cdot}1$, $\mathrm{L}(B)$ is about 10% distant from B and an error of less than 0·01% in its value is unimportant. Hence the third term may be neglected and it is quite adequate to write

$$\frac{\mathrm{L}(B)}{B} = \frac{1 - \frac{K^2t^2}{B^2+1}}{1 - Kt}$$

The approximation will be even better for smaller values of Kt.

Similarly an approximate expression for the deviation of the limits from the best estimate is

$$\frac{\mathrm{L}(B) - B}{B} = Kt\left\{\frac{1 - \frac{Kt}{B^2+1}}{1 - Kt}\right\}$$

(*d*) *Transformation of the Functional Relationship Equation*

The usual form of the equation is

$$b = \frac{s_{yy} - \lambda s_{xx}}{2s_{xy}} + \sqrt{\left[\left(\frac{s_{yy} - \lambda s_{xx}}{2s_{xy}}\right)^2 + \lambda\right]}$$

when s_{xy} is positive.

Let $$g^2 = \frac{s_{xx}}{s_{yy}}$$

Then $$r = \frac{s_{xy}}{\sqrt{(s_{xx}s_{yy})}} = \frac{s_{xy}}{gs_{yy}}$$

$$\therefore \quad \frac{s_{yy}}{s_{xy}} = \frac{1}{gr}$$

$$b = \frac{s_{yy}(1-\lambda g^2)}{2s_{xy}} + \sqrt{\left[\left(\frac{s_{yy}(1-\lambda g^2)}{2s_{xy}}\right)^2 + \lambda\right]}$$

$$= \frac{1-\lambda g^2}{2gr} + \sqrt{\left[\left(\frac{1-\lambda g^2}{2gr}\right)^2 + \lambda\right]}$$

Let $$C^2 = \frac{\mathrm{V}(y)/s_{yy}}{\mathrm{V}(x)/s_{xx}}$$

Now $$\lambda = \frac{\mathrm{V}(y)}{\mathrm{V}(x)}$$

$$\therefore \quad C^2 = \lambda g^2$$

$$b = \frac{1-C^2}{2gr} + \sqrt{\left[\left(\frac{1-C^2}{2gr}\right)^2 + \lambda\right]}$$

If $\bar{b}$ is the value when $C=1$, i.e. when $\lambda = 1/g^2$

Then $$\bar{b} = 1/g$$

$$\frac{b}{\bar{b}} = \frac{1-C^2}{2r} + \sqrt{\left[\left(\frac{1-C^2}{2r}\right)^2 + C^2\right]}$$

REFERENCES

[1] Williams, E. J. (1959). *Regression Analysis*. John Wiley and Sons, New York; Chapman and Hall, London.

[2] Lindley, D. V. (1947). 'Regression lines and the linear functional relationship', *Suppl. J. R. Statist. Soc.*, **9**, 218.

[3] Creasy, Monica A. (1956). 'Confidence limits for the gradient in the linear functional relationship', *J. R. Statist. Soc.*, B, **18**, 65.

AN INTRODUCTION TO DYNAMIC PROGRAMMING*

M. G. SIMPSON

In the course of their work statisticians are frequently called on to make the acquaintance of, and possibly to use, techniques from other fields, such as operational research. One such technique is Dynamic Programming; in this expository article Dr Simpson explains it in simple terms.

THIS paper is mainly intended to map out the general field of application of dynamic programming and to illustrate its basic concept, but the computation required is also indicated.

I must first make it clear that dynamic programming has little in common with the better-known technique of linear programming. It is true that there are many problems which may be solved using either technique, but in such cases it is nearly always preferable to use linear programming with its much simpler computation. Dynamic programming may, however, be applied to a wide range of problems to which linear programming is not applicable; other advantages of using the technique will be indicated later.

Multi-Stage Decision Processes

The technique was devised initially (Richard Bellman, *Dynamic Programming*, Oxford University Press, 1957) for the solution of a class of problems called multi-stage decision processes. By this we mean processes where decisions are taken sequentially and where they are not independent, in that decisions taken early on in the process affect decisions made subsequently. For example, if we consider a retailer who restocks with a certain commodity at, say, monthly intervals, then the amount which he orders in one month will certainly affect the amount he orders in the succeeding months. For if he orders too large an amount of stock he will probably require to order little if any the following month, whereas if he places too small an order initially he will run out of the commodity during the month and will have to order a substantial amount next time. In this we can see that the decisions made affect decisions to be made subsequently. This is the most straightforward class of multi-stage decision processes, in which the decisions are made sequentially in time. Although processes of this type arise sufficiently frequently for a technique for handling them to be worth while, it is also possible to tackle other problems in a similar manner. In certain problems, even if the time element is not explicit,

* Based on a paper read at a one-day conference on Dynamic Programming organised by the Birmingham Group of the Royal Statistical Society, May 1960.

it is sometimes advantageous to treat them as multi-stage processes and to use dynamic programming to solve them. An example which we shall consider in some detail is of this nature.

Characteristics of Multi-Stage Decision Processes

The types of processes which we can tackle using dynamic programming can be sub-divided in several ways.

For instance we can deal with problems which are either deterministic or stochastic. In deterministic processes the result of any decision is unique and is known beforehand. Stochastic processes are those in which the result of any decision is not known uniquely but can only be forecast in terms of some probability distribution—the retailer mentioned above is usually in this unfortunate stochastic state.

We can also consider problems in which the number of decisions available at each stage in the process is either finite or infinite. A 'finite' problem may have only two decisions possible at any stage, whereas an 'infinite' problem might arise when we have to divide some resource in a continuous manner at each stage.

The final main property we have to consider is the number of stages in the process. In the simplest case the number of stages is finite and known beforehand. There are problems, however, in which the number of stages is infinite. Such problems can arise either because we have a process in which decisions are taken sequentially in time, and we wish to continue the process indefinitely, or because decisions are being taken continuously over a given range or period, and the stages become infinitesimally small. Finally there are certain problems in which the number of stages is not known beforehand, but has to be determined in the solution.

We have, therefore, three main properties in the problems we can tackle using dynamic programming: whether the process is deterministic or not, the number of decisions at each stage in the process, and the number of stages. Fortunately, however, the formulation of problems by dynamic programming is very similar whatever the characteristics of the problem.

Two Typical Problems

To complete this outline of the field of application of the technique, let us briefly consider some typical problems.

The simplest type is one in which we have a deterministic process, with a given number of stages, and a limited number of decisions at each stage. An example of this type might be a cargo loading problem, in which we have a variety of cargoes, with different weights per unit, and different values per unit weight, and we wish to load a certain capacity with a cargo of maximum value. In this problem time is not explicit, but we can consider it as a multi-stage process, where each stage constitutes the decision on how many of each type of cargo to load. The problem is clearly deterministic in that the values and weights are

known exactly, the number of stages is known, and the number of possible decisions at each stage, though not known, is at least made finite by the capacity available.

A more complex problem is the following one of replacement. Suppose we have a set of machines with variable lives, but each life expectancy is known as a distribution of times from previous history. In any month, say, we can either replace a machine at some cost, or let it continue in service, with the chance of incurring substantial costs if it breaks down. We want to know what replacement policy leads to minimum cost. Such a problem is essentially stochastic, in that machine life is variable. The number of possible decisions at each stage is finite (in fact only two for a single machine—to replace it or not), but the number of stages is infinite if we wish to continue in the same line of business.

The Principle of Optimality

The way in which we tackle problems of this type is by applying, by sometimes devious means it is true, a very simple concept. The concept was stated by Bellman as the Principle of Optimality, in the following terms: 'An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.' This might seem rather formidable, but in essence it is equivalent to saying that if we are pursuing an optimal policy, then at each stage during the process the remaining decisions must themselves form an optimal policy from the state reached to the end-point of the process.

This is, of course, almost a truism and a proof by contradiction is immediate. For all its simplicity, however, it is very powerful in providing a means of tackling what are otherwise intractable or very difficult problems. For it enables us to relate the solution of a problem of four stages, say, to a similar problem containing five stages, and in fact to relate an n-stage problem to a $(n+1)$-stage problem. In this way we can establish a series of recurrence relations; once we can find a suitable starting-point (say the one-stage problem), we can extend this solution step by step to problems involving any number of stages.

A Simple Application of the Principle of Optimality

Before we tackle a problem in some detail, let us look briefly at a simple outline of the procedure.

Suppose we wish to travel from a town A to a town Z, and in doing so we have to pass through a series of intermediate towns, (B) to (Y). One town in each of the sets (B) to (Y) must be included, but we have a choice of routes through these. We want to find the minimum-cost route from A to Z, knowing the route costs between all towns.

The way in which we can use the Principle of Optimality in this sort of problem is as follows:

Consider first all the towns in group (B). These have associated with them costs from A, say $\alpha(B_i)$ for all i, and $\alpha(B_i)$ is the minimum (in fact here the only) cost of reaching B_i from A.

Now consider any town (C). The cost of reaching C_i depends on the intermediate town in (B) visited, but for each C_i there will be at least one of (B) giving a minimum-cost route to C_i. Thus for each C_i we have to scan all (B) to find the one giving minimum cost, say $\alpha(C_i)$. This is all straightforward so far, but when we come to the next stage, we find the minimum cost of reaching D_i by scanning all the (C) to find the one which has minimum sum $\alpha(C_j) + \text{cost } (C_j, D_i)$, i.e. the minimum-cost route to D_i includes as part of it the minimum-cost route to C_j, and we have only to scan towns (C), and not concern ourselves explicitly with towns (B). Thus we obtain the minimum-cost routes to (D) from the minimum-cost routes to (C) and have only one variable [which of the towns (C) to include] at this stage.

This procedure can be followed through until the last stage is reached, when finally the minimum-cost route from A to Z is obtained by scanning only towns (Y) to find which of these has the minimum sum $\alpha(Y_j) + \text{cost } (Y_j, Z)$. Every time we include an extra step in the process we are applying the Principle of Optimality.

Thus we are able to proceed logically step by step through the problem, and although with only four towns at each stage, and all towns in successive stages connected, there is a very large number of possible routes from A to Z, we can find the minimum-cost route by setting up 25 successive functions, each having a value at only four points and very easily computed.

This is the sort of procedure we follow in all dynamic programming problems. Let us now move on to a more practical problem in which we shall set up the type of function used in dynamic programming and indicate the computational steps required.

A Multi-Activity Investment Problem

Suppose we have a sum of money to be invested in various activities, with the relation between investment and the return known for each activity—i.e. activity i has a return $f_i(x)$ from an investment x.

We require to know how to allocate the total sum available, S, among the activities so as to obtain the maximum total return.

If there are n activities, we wish to determine that set of (x_i), where i goes from 1 to n, so that $\Sigma x_i = S$, and $\Sigma f_i(x_i)$ is maximised.

For certain types of functions, there are simple ways of solving this problem, but there are forms for $f_i(x)$ which will involve using successive approximation methods to obtain the solution. However, if n is large, such methods are very laborious, and can often be dangerous in arriving at local, but not absolute, maxima.

Dynamic programming can be used whatever the form of the functions, and though there is still a certain amount of computation required, it is much less than will usually be required in an n-variable

iteration, and there is no danger of finishing at a local maximum.

We can tackle the problem by defining a function $R_i(x)$ as being the maximum return obtainable from an investment sum x, when we allow investment only in activities 1 to i inclusive. Thus $R_1(x) = f_1(x)$ as long as $f_1(x)$ is an increasing function.

Consider now investment in activities 1 and 2; if $r_2(x, y_2)$ is the return from a total amount x when y_2 is invested in activity 2, we have

$$r_2(x, y_2) = f_1(x - y_2) + f_2(y_2)$$

But if we wish to maximise the return from our x, we must choose y_2, for the given x, so that the joint return is maximised, i.e.

$$R_2(x) = \text{Max}[r_2(x, y_2)]$$

and since y_2 may have any value between 0 and x, and since from the Principle of Optimality we must also use the amount invested in activity 1 in the optimal fashion, we have

$$R_2(x) = \underset{0 \leqslant y_2 \leqslant x}{\text{Max}} [R_1(x - y_2) + f_2(y_2)]$$

For each value of x we can find a value y_2, dependent on x, which maximises this expression. The maximum may, if the form of $f_1(x)$ is suitable, be found by differentiation, but if we do use calculus, we must ensure that we obtain the *absolute* maximum, rather than some turning value which is exceeded at other points in the range.

Of course, if we use numerical methods to obtain the maximum, we can handle any functions $f_i(x)$. In any event, we are able to compute $R_2(x)$ for all x in the range $(0, S)$, and this function gives us the maximum return obtainable from a sum x, when we permit investment in activities 1 and 2 only. (The order in which the activities are taken is unimportant, but it simplifies the notation if we take them in numerical sequence.)

The way in which the sum x is divided between the two activities depends of course on the particular functions f_1 and f_2, and for some forms the maximum return will be obtained if the whole sum x is put into one or other of the activities. The function $R_2(x)$ does not explicitly reveal how the sum is divided, but only gives the return from making the division in the best fashion. We shall, however, need to know how to make this division optimally in order to know the best policy to adopt. The information is already at hand, of course, since we have determined, for each value of x, that part y_2 which is put into activity 2 in order to obtain the maximum return.

If we denote $y_2 = y_2(x)$ as the value of y_2 which gives the maximum return from a total x, then we can regard $y_2(x)$ as a decision function, telling us how much we should invest in activity 2 when x is available for the activities 1 and 2 jointly. We now have all the information we require for the 2-stage process, and can use this to extend the solution to the 3-stage process, and sequentially to an n-stage process.

For if we now permit investment in activity 3, investing y_3 in this from a total x available, we have the return as

$$r_3(x, y_3) = f_3(y_3) + \text{return from } (x - y_3) \text{ in activities 1 and 2.}$$

But if we wish to maximise the total return, again using the Principle of Optimality, we must utilise the $(x - y_3)$ in an optimal fashion with respect to activities 1 and 2; therefore

$$r_3(x, y_3) = f_3(y_3) + R_2(x - y_3)$$

since we defined $R_2(x - y_3)$ as being the maximum return obtainable from a sum $(x - y_3)$ with investment in activities 1 and 2.

We still have to ensure that the y_3 is chosen optimally, and we do this as before, by letting it take all possible values (i.e. in the range 0, x) and choosing that particular y_3 which maximises the return, i.e. we compute

$$R_3(x) = \underset{0 \leqslant y_3 \leqslant x}{\text{Max}} [f_3(y_3) + R_2(x - y_3)]$$

Again this is easily obtained, being a single variable maximisation. This gives us the maximum return from three activities, and we may again compute this for all x in the range $(0, S)$. At the same time we automatically generate the corresponding decision function $y_3(x)$, denoting the amount y_3 which is invested in activity 3 when x is available for activities 1, 2, and 3.

We can proceed in this manner, since for all $i \geq 2$ we have

$$R_i(x) = \underset{0 \leqslant y_i \leqslant x}{\text{Max}} [R_{i-1}(x - y_i) + f_i(y_i)]$$

as the return function and $y_i = y_i(x)$ as the corresponding decision function.

Finally we have

$$R_n(x) = \underset{0 \leqslant y_n \leqslant x}{\text{Max}} [R_{n-1}(x - y_n) + f_n(y_n)]$$

for all x in $(0, S)$ as the maximum return obtainable from all n investments, and in particular

$$R_n(S) = \underset{0 \leqslant y_n \leqslant S}{\text{Max}} [R_{n-1}(S - y_n) + f_n(y_n)]$$

as the maximum return with our particular sum S available.

This tells us part of the answer, but to determine how this maximum return is obtained we have to trace back through the problem using the successive decision functions in the reverse order.

For $y_n(x)$ tells us how much we should invest in activity n, when we have x available for all activities, so that $x_n = y_n(S)$ is how much should be invested in the nth activity for our particular problem. We should note that it is only at this final stage that we commit ourselves to any particular policy, and it is this which ensures that we obtain the absolute maximum return, rather than some local maximum.

Having obtained x_n, we have $(S - x_n)$ left to invest in the remaining $(n-1)$ activities, and the function $y_{n-1}(x)$ tells us that we put $x_{n-1} = y_{n-1}(S - x_n)$ into the $(n-1)$th activity.

Similarly $x_{n-2} = y_{n-2}(S - x_n - x_{n-1})$ and we proceed back through the problem to find how much to invest in each activity.

This concludes the computation required in the problem, and though we still have an appreciable amount of work, we have reduced the labour significantly by replacing the original n-variable problem by n simple one-variable maximisations.

The Field of Solutions

With regard to the investment problem discussed in the previous section, our analysis provided far more information than the original problem required. We have, in fact, the solutions to a whole set of problems similar in form to the original, but with different parameters.

For our functions $R_n(x)$ give us the maximum return from all sums from 0 to S, and our intermediate functions $R_i(x)$, the maximum return from activities 1 to i inclusive from sums of 0 to S. This additional by-product information is typical of dynamic programming, for inherent in the application of this technique is the necessity to embed the problem in a whole field of similarly structured problems.

In an industrial environment we often require solutions to a field of problems, in order to determine what policy should be followed in a range of circumstances.

Sometimes the need to generate solutions in this way can avoid a lot of the detailed computation by giving an idea of the form of the solution. It is often possible, for instance, having defined the series of functions of which one relates to the specific problem under consideration, to see from the first few stages exactly what is the structure of the optimal policy and to infer the form of the final solution.

This procedure is helped in some problems by our having two parallel sets of functions—the return function and the decision function, for we can use either of these to help us to get an insight into the real structure of the problem.

DYNAMIC PROGRAMMING AND DECISION THEORY*

D. V. LINDLEY†
Statistical Laboratory, University of Cambridge

The previous article in this issue dealt with the general principles of Dynamic Programming. In this article Professor Lindley shows how Dynamic Programming links up with certain decision problems in the statistical field.

Introduction

The subject of this paper is the contribution of dynamic programming to some sequential statistical (or decision) problems. The first, and the most familiar, truly sequential procedure is Wald's[7] likelihood ratio test; this was originally designed to test one simple hypothesis against another, but is often used to deal with composite hypotheses. It is remarkable that although Wald's test is about twenty years old it is almost the only sequential procedure which is known to be *optimum* in some satisfactory sense (Wald and Wolfowitz[9]): for example, it is not known what is the optimum procedure for deciding between three (as distinct from two) simple hypotheses. This is not to say that these twenty years have not seen the introduction of useful sequential procedures, they have; my point is that these are not known to be best in any useful sense. In this paper I hope to show that dynamic programming provides a *computational* technique for finding the *optimum* sequential procedure in a wide variety of statistical situations. Examples of useful, if not necessarily optimum, procedures are the sequential t-test (see, for example, National Bureau of Standards[6]); applications for medical purposes devised by Armitage[1]; and the procedures for finding the best operating conditions, due to Box and Wilson.[3] In this paper I shall discuss, in particular, the problem of deciding whether the unknown mean of a normal distribution of known variance is positive or negative—a seemingly simple problem which is not completely solved. But before I do this it is necessary to make some remarks about decision problems, particularly in reference to Bayesian ideas.

Non-sequential Decision Problems

Suppose that there is some 'unknown' state of nature denoted by θ. Usually θ will consist of a single real parameter (such as a mean or variance) or a set of real parameters. Suppose further that it is possible to make an observation, resulting in the value x, and that the probability distribution of the observation depends on θ in a known way.

* Based on a paper read at a one-day conference on Dynamic Programming organised by the Birmingham Group of the Royal Statistical Society, May 1960.

† Present address: University College of Wales, Aberystwyth.

Denote the probability density of x, given θ, by $p(x|\theta)$; a known function. So far this is the usual statistical situation. But now let $\pi(\theta)$ be a prior probability distribution for θ supposed to represent the statistician's beliefs about θ before making the observation. For example, if he were asked whether θ were positive or negative he would say that the odds were

$$\int_{\theta \geqslant 0} \pi(\theta)d\theta \qquad \text{to} \qquad \int_{\theta < 0} \pi(\theta)d\theta$$

that it was positive. It is not the place here to discuss the need for such a probability distribution over the states of nature; it is enough to say that its introduction is basic to dynamic programming ideas in sequential experimentation, and that it describes our knowledge of θ. No state of nature is truly 'unknown' in the sense used by most contemporary statisticians.

Now consider what happens to our knowledge of θ when the observation is made: $\pi(\theta)$ changes according to Bayes's theorem; the posterior distribution of θ, given x, is

$$\pi(\theta|x) = p(x|\theta)\pi(\theta)/p(x) \qquad \ldots . (1)$$

where

$$p(x) = \int p(x|\theta)\pi(\theta)d\theta \qquad \ldots . (2)$$

That is, the posterior distribution of θ is proportional to the product of the known density of x and the prior distribution of θ. The constant of proportionality is the inverse of $p(x)$, which is itself of interest, because it is the prior distribution of the observation before it is made. If the statistician were asked his views about what value the observation would assume, then he would express them through $p(x)$. We can now see the role that observations play: they change the distributions over the states of nature (from prior to posterior) and, for example, it can be shown that, on the average, the distributions become, in an obvious intuitive sense, more concentrated; thus the distributions reflect an average gain in knowledge, for a concentrated distribution means that the value of θ is more accurately known than when the distribution has a large spread. The appreciation of this change in the distribution of θ with observations is important in understanding the sequential ideas below.

Finally suppose it is necessary to take one amongst a number of decisions on the basis of one's knowledge of the state of nature. If θ were known, presumably the best decision d would be known. In default of this knowledge it is necessary to introduce a utility function $U(d, \theta)$ describing the utility of taking decision d when the true state of nature is θ. For a given θ the best decision is that which maximises the utility $U(d, \theta)$. Now after the observation is made, when the knowledge of θ is given by $\pi(\theta|x)$, the best decision to take is that which maximises the *expected* utility, the expectation being taken over θ, for the true

utility is unknown since it depends on θ. Hence take that d which maximises

$$\mathrm{E}_\theta\{\mathrm{U}(d, \theta)\} = \int \mathrm{U}(d, \theta)\pi(\theta|x)\mathrm{d}\theta = \mathrm{U}^*(d, x) \quad \ldots . (3)$$

say. Had the observation not been available then $\pi(\theta)$ would have had to replace $\pi(\theta|x)$ and the expectation would have been

$$\int \mathrm{U}(d, \theta)\pi(\theta)\mathrm{d}\theta = \mathrm{U}^*(d)$$

say. In either case we shall take the decision which maximises the utility: with an observation this will yield $\max_d\{\mathrm{U}^*(d, x)\} = \bar{\mathrm{U}}(x)$, say, whereas without it the yield will be $\max_d\{\mathrm{U}^*(d)\} = U_0$, say. Now $\bar{\mathrm{U}}(x)$ may exceed or be less than U_0, but on the average it exceeds U_0. By 'on the average' here we mean an average over the possible observations that might arise; some observations may reduce the utility but the average effect is an increase. Now before the observation is made our prior distribution for the observation is $\mathrm{p}(x)$, equation (2), so that the average just referred to is

$$\mathrm{E}_x\{\bar{\mathrm{U}}(x)\} = \int \bar{\mathrm{U}}(x)\mathrm{p}(x)\mathrm{d}x = \bar{U} \quad \ldots . (4)$$

say. It is easy to prove that $\bar{U} \geq U_0$ and the difference $\bar{U} - U_0$ gives the expected gain in utility provided by the observation. The observation is only worth making if this expected gain (which is necessarily non-negative) exceeds the cost (measured in the same units as utility) of the observation. An example may clarify the situation.

*Example of a Non-sequential Decision Problem**

Suppose that θ is the mean of a normal distribution of variance unity and suppose that x is the mean of a random sample of n observations from this distribution. (Any other function of the observations besides the mean is irrelevant; in technical language the mean is sufficient.) Then

$$\mathrm{p}(x|\theta) = \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}n(x-\theta)^2\right] \quad \ldots . (5)$$

Suppose that $\pi(\theta)$ is normal with mean θ_0 and variance τ^2, we write $\mathrm{N}(\theta_0, \tau^2)$; and that there are two decisions d_1 and d_2 with

$$\left.\begin{aligned} \mathrm{U}(d_1, \theta) &= c_1 \\ \text{and} \quad \mathrm{U}(d_2, \theta) &= c_1 + b(\theta - \bar{\theta}) \end{aligned}\right\} \quad \ldots . (6)$$

where $b > 0$. Then if θ is known, d_2 is the preferred decision if $\theta > \bar{\theta}$, otherwise d_1 is preferred. In other words we have to decide whether θ is greater or less than $\bar{\theta}$, the utility of deciding that it is greater than $\bar{\theta}$ increasing linearly with θ.

* This example was discussed in an unpublished paper which I gave to the Loughborough conference of the Royal Statistical Society in September 1959.

The calculations in this problem are much simplified by noting the following fact: the posterior distribution of θ is still normal. This is a consequence of simple calculations carried out with Bayes's formula, equation (1). The mean and variance of the posterior distribution, for those who care to check the calculations, are respectively

$$\frac{xn+\theta_0/\tau^2}{n+1/\tau^2} \quad \text{and} \quad [n+1/\tau^2]^{-1} \qquad \ldots . (7)$$

Notice that the mean is a weighted average of x, the mean of the observations, and θ_0, the prior mean; and that the weights are, in the usual way, inversely proportional to the variances. In other words we can think of normal observations as changing the mean of the distribution of θ in a *random* way but changing the variance *deterministically*, retaining a normal distribution for θ.

From this fact it is easy to verify that d_1 is only taken when

$$x < \bar{\theta} + (\bar{\theta} - \theta_0)/n\tau^2 \qquad \ldots . (8)$$

If $\bar{\theta} = \theta_0$ this means that d_1 is preferred only when the observed mean is less than the critical value. But when $\bar{\theta} \neq \theta_0$ the additional term $(\bar{\theta} - \theta_0)/n\tau^2$ appears and allows for the prior belief, decreasing as n or τ^2 increase. It is also possible, using some rather more tedious calculations (given earlier by Grundy *et al.*[5]) to evaluate $\bar{U}$ [equation (4)] and hence the expected value of the observations, $\bar{U} - U_0$. If the cost per observation is c then n observations are worth making only if $\bar{U} - U_0 > cn$. The optimum number of observations is the value of n which maximises $\bar{U} - U_0 - cn$, the gain in utility expected from the experiment. In a sense such a procedure for deciding the number of observations is sequential, but it only becomes truly sequential when the observations are taken one at a time and at each stage the experimenter decides whether or not to take another one. To obtain the optimum procedure here involves new difficulties and we now consider these.

Sequential Decision Problems

Each stage of a truly sequential problem, with observations taken one at a time, is a decision problem of the type considered on p. 40, with the number of decisions increased by one by including the decision 'go on and take another observation'. The same remarks therefore apply to sequential problems as to non-sequential ones, with the addition that the utility of this extra decision has to be found. Now it is easy (by the methods sketched above) to find the expected utilities of stopping and taking any particular decision, but it is not easy to find the expected utility of taking one more observation. For the expected utility required is that of going on and then doing the best possible from then onwards. Consequently in order to find the best decision *now* (i.e. in particular whether to go on or not) it is necessary to know the best decision *in the future*. In other words the natural time order of working from the present to the future is not of any use because the present

optimum involves the future optimum. The only method is to work backwards in time: from the optimum future behaviour to deduce the optimum present behaviour, and so on back into the past. It might be thought to be enough to consider the consequences of either stopping or taking just one further observation and then stopping; but this is not adequate because there are common situations in which, for example, two observations are worth having, but one is not. The *whole* of the future must be considered in deciding whether to go on.

For example in the problem of the normal mean on p. 41, where the observations are taken one at a time in order to decide whether θ is less than or greater than $\bar{\theta}$, it is intuitively clear that the optimum procedure will be a member of the following class: Let $\{a_n\}$, $\{b_n\}$ be two sequences of constants, with $a_n \leqslant b_n$, and let x_n be the mean of the first n observations; if $x_n < a_n$, take decision d_1; if $x_n > b_n$, take decision d_2; and if $a_n \leqslant x_n \leqslant b_n$ go on and take a further observation. It is impossible to determine, say, a_8 and b_8, until the two sequences are known for $n > 8$; because, until they are, the expected utility of going on when $n = 8$ cannot be evaluated and compared with the expected utilities of stopping and deciding finally whether θ is less than or greater than θ_0. Consequently we have to seek for an expression for a_8 and b_8 in terms of the sequences for $n > 8$. Even this is difficult, and it is here that dynamic programming ideas help. But before considering them let me repeat one remark made on p. 42: at any stage of the sequential procedure considered there the position may be summarised by two quantities, the mean and the variance of the distribution of θ at that stage. Furthermore the variance changes deterministically whilst the mean changes randomly according to a known probability law: thus one observation will change $N(\theta_0, \tau^2)$ into $N(\theta_1, [1 + 1/\tau^2]^{-1})$ where

$$\theta_1 = \frac{x + \theta_0/\tau^2}{1 + 1/\tau^2}$$

In practice it is easier to work with the inverse of the variance because it changes additively, namely from τ^{-2} to $1 + \tau^{-2}$. The distribution of θ_1 for given θ_0 is known since the distribution, $p(x)$, of x is known [equation (2)].

Dynamic Programming

In the previous article in this issue Dr Simpson described, in general terms, the basic important idea behind dynamic programming. In decision theory problems it is the idea of concentrating attention, not on the optimum procedure itself, but on the *utility* expected from the optimum procedure. Once the optimum utility has been found it is usually a simple problem to find that procedure which produces this utility. What we do in applying dynamic programming ideas to decision problems is to write down an equation for the optimum utility in the hope of solving it numerically or analytically.

The expected optimum utility at any stage is, by the discussion on

pp. 40-41, a function of the distribution of θ at that stage, and, by the remark repeated at the end of the last section, this distribution depends, in many cases, on two parameters. Let $U(w, t)$ be the expected utility of the *best sequential* scheme when starting from a situation in which the parameters describing the distribution of θ are w and t; w changes randomly and t deterministically. In the normal mean situation w is the mean of the distribution of θ and t is the inverse of the variance. Let $\bar{U}(w, t)$ be the expected utility of taking the best decision now, without any further observations; $\bar{U}(w, t)$ is a known function. Then $U(w, t)$ is either $\bar{U}(w, t)$ (if it is not worth taking further observations) or can be evaluated in terms of the expected utility after taking a further observation (if an observation is taken). In the latter case w and t change to, say w' (randomly) and $t+h$ (deterministically), h is one in the example. Let $p(w'|w, t)$ be the distribution of w' given w and t: as explained above this is a known function. Lastly let $c(t, h)$ be the cost of increasing t to $t+h$. Then it is clear that if experimentation is continued

$$U(w, t) = \int U(w', t+h)p(w'|w, t)dw' - c(t, h) \quad \ldots . (9)$$

Consequently the basic equation of dynamic programming in this sequential decision problem is

$$U(w, t) = \max\{\bar{U}(w, t): \int U(w', t+h)p(w'|w, t)dw' - c(t, h)\} \ldots . (10)$$

in which everything is known except $U(w, t)$. If it can be solved for $U(w, t)$ then the optimum procedure is obvious: go on only when $U(w, t) > \bar{U}(w, t)$; otherwise stop and take the optimum decision obtained as in the non-sequential case. This illustrates the remark made earlier that once the optimum utility has been found it is easy to find the procedure which produces this utility.

The most important point to notice about equation (10) is that it is well suited for numerical work. If $U(w', t+h)$ is known for all w' and some single value of $t+h$ then, from (10), it is a simple matter of numerical integration and comparison with $\bar{U}(w, t)$ to calculate $U(w, t)$ for all w and the single value t. In the normal example, and typically, t corresponds to the sample size; hence one can work backwards in order of *decreasing* sample size as explained above. If the scheme is a truncated one, the point of truncation provides a convenient starting point. For example, in Armitage's medical work, it would be easy to work out for any particular case the best truncated scheme for deciding between three simple alternatives.

Satisfactory as the position is with regard to particular numerical cases, the theory and analytic solution of equation (10) are practically non-existent. I had hoped to be able to report the solution to the normal case but it has eluded me. [In the interval between delivering the lecture (May 1960) and writing it up for publication (September 1960) Chernoff[4] gave a paper in which he reached almost the same position as myself. In a personal communication he says that he hopes

to perform the necessary numerical calculations.] Despite this lack of success it may be worth describing the position in that simple case. Equation (10) involves the variable w continuously whereas the variable t occurs discretely in steps of amount h. It would simplify the situation if the dependence on t could also be continuous (on the general grounds that differentials are easier to handle than differences). Now t is the inverse of the variance and so [from equation (7)] is equal to $n+\tau^{-2}$. Had each observation had variance σ^2 instead of unity, it would have been $n/\sigma^2+\tau^{-2}$. Now allow $\sigma^2 \to \infty$ and $n \to \infty$ in such a way that n/σ^2 remains constant; then t can assume a continuous range of values. What we are doing here is to say that one observation of variance unity is equivalent to n observations each of variance n and supposing that we are taking a large (and ultimately infinite) number of observations each of large (and ultimately infinite) variance. The discontinuities caused by taking an observation are therefore removed. In this limit it is not difficult to show that the *integral* equation (9), which holds whenever it is worth taking another observation, is replaced by the familiar diffusion *differential* equation whose solution, in terms of the normal density function, is well known for given boundary conditions. Consequently in the region in the (w, t)-plane in which sampling continues the diffusion equation obtains for $U(w, t)$ and outside this region $U(w, t)$ is equal to the known function $\bar{U}(w, t)$. Furthermore the function $U(w, t)$ must be continuous on the boundary between these two regions and it is the boundary, not the function itself, that is of principal interest. Consequently we have a free-boundary problem in which the boundary is the main unknown. The natural way, therefore, to transform (10) is to derive from it an equation for the boundary. At first this does not seem difficult, but prolonged investigation has not, so far, enabled me to resolve the difficulties. I conjecture that the boundary [in the (w, t)-plane] satisfies a differential equation which is of the first order and second degree.

A general point I wish to emphasise here is that dynamic programming is, at its present stage of development, a technique which is suitable for numerical work, but that little is known about the solution of the equations, like (10), or about the properties of the solutions. It may sound old-fashioned in these days of fast machines to decry the solution of particular problems by their use; but it does seem to me that we should gain much more understanding of the problem by solving the equation analytically rather than numerically. It is true, in the normal case, that by transformations all the parameters can be removed so that a single boundary would suffice for all parameters; nevertheless a numerical solution of the differential equation—if we knew it—would be simpler than the iteration of (10) directly.

A word may be inserted here about the origin of equation (10). The phrase dynamic programming and the central ideas associated with it are obviously due to Bellman[2] but the approach had been used earlier in sequential decision problems by Wald.[8] Theorem 4.2 on page 105

of Wald's book is essentially our equation (10). I am indebted to Professor Barnard for this reference.

*Sufficiency**

In this section I show that the expected utility depends on two parameters, w and t, w changing randomly and t deterministically, in a wide class of situations. Suppose that an observation depends on a parameter θ consisting of s real numbers $(\theta_1, \theta_2, \ldots \theta_s)$, with a probability density of the following form

$$p(x|\theta) = F(x)G(\theta) \exp\left[\sum_{i=1}^{s} f_i(x)\theta_i\right] \quad \ldots\ldots (11)$$

i.e. a product of a function of the observation only, a function of the parameter only, and an exponential term linear in the components θ_i of the parameter. Notice that G is determined by F since the integral of (11) over x must be unity: conversely G determines F. If the prior distribution of θ is $\pi(\theta)$ the posterior distribution is proportional to

$$G(\theta)\pi(\theta) \exp\left[\sum_{i=1}^{s} f_i(x)\theta_i\right] \quad \ldots\ldots (12)$$

the constant of proportionality depending only on x, and being determined from the condition that the integral be unity. Now suppose n independent observations $(x_1, x_2, \ldots x_n)$ from (11) are taken. The joint density is the product of n terms like (11) and the posterior distribution $\pi(\theta|x_1, x_2, \ldots x_n)$ is proportional to

$$G(\theta)^n\pi(\theta) \exp\left[\sum_{j=1}^{n}\sum_{i=1}^{s} f_i(x_j)\theta_i\right] \quad \ldots\ldots (13)$$

Now this distribution of θ is characterised by $(s+1)$ parameters: the s coefficients of the θ_i, namely

$$w_i = \sum_{j=1}^{n} f_i(x_j)$$

and n, the sample size. Consequently at any stage in the sequential scheme the distribution of θ can be described in terms of $w = (w_1, w_2, \ldots w_s)$ and $t = n$. Hence the expected utility is also a function of w and t. Also w changes randomly with the addition of a further observation—in fact $w_i' = w_i + f_i(x_{n+1})$ so that it is an additive random walk—and t changes deterministically. The form (11) is that of the exponential family, the family of distributions which possess s sufficient statistics

$$\left[\sum_{j=1}^{n} f_i(x_j)\right]$$

for any size of sample. Most common distributions belong to this family: for example the normal, Poisson, binomial, and gamma distri-

* This section was not included in the conference paper, for lack of time, but is given here for completeness.

butions. (J. A. Lechner, in an unpublished Ph.D. thesis at Princeton University, has discussed the same example as ours with the substitution of the Poisson process for the normal one.)

The Marriage Problem

I conclude this paper with a very simple sequential decision problem for which the basic equation can be solved very easily. It should not be taken too seriously, and lady readers should interchange the sexes in the text. A known number, n, of ladies are presented to you one at a time in a random order. After inspecting any number r $(1 \leqslant r \leqslant n)$ of them you are able to rank them from best to worst and this order will not be changed if the $(r+1)$th lady is inspected; she will merely be inserted into the order. At any stage of the 'game' you may either propose to the lady *then being inspected* (there is no going back!), when the game stops, or inspect the next lady; however, if you reach the last lady you have to propose. All proposals are accepted. What is the optimum strategy? Before this has a definite answer we must assign a utility function: let U_i be the utility of being married to the lady with the ith true rank, so that $U_i \geq U_{i+1}$. True rank is the rank she has amongst all n ladies, as distinct from the apparent rank that she has when only some have been inspected. We can again argue in terms of two co-ordinates: r, the number of ladies inspected at any stage and s, the apparent rank of the rth lady being inspected. Now r changes deterministically in steps of one and s changes randomly; so r plays the role of t and s that of w in the above discussion.

In the notation used in our basic equation (10), t is the variable corresponding to the number of ladies inspected at any stage, here r; w is the variable corresponding to the rank of the lady now being inspected, here s. So $\mathrm{U}(w, t)$ is replaced by $\mathrm{U}(s, r)$ and $\bar{\mathrm{U}}(w, t)$ by $\bar{\mathrm{U}}(s, r)$. The probability that the rth lady of apparent rank s will have true rank S is easily calculated as

$$\binom{S-1}{s-1}\binom{n-S}{r-s} \Big/ \binom{n}{r} = p_{S:s,r} \qquad \ldots . (14)$$

say, so that

$$\bar{\mathrm{U}}(w, t) = \bar{\mathrm{U}}(s, r) = \sum_{i=s}^{s+n-r} U_i p_{S:s,r} \qquad \ldots . (15)$$

Given the situation described by r and s, the probability that the next lady will have apparent rank s' is clearly $(r+1)^{-1}$ for all s', so that $\mathrm{p}(w'|w, t)$ in (10), now $\mathrm{p}(s'|s, r)$, is $(r+1)^{-1}$. Hence (10) is

$$\mathrm{U}(s, r) = \max\left\{\bar{\mathrm{U}}(s, r): \sum_{i=1}^{r+1} \mathrm{U}(i, r+1)/(r+1)\right\} \qquad \ldots . (16)$$

with $\bar{\mathrm{U}}(s, r)$ a known function given by (15). There is no cost of inspection.

We consider two special cases. In the first $U_1=1$ and $U_i=0$ for $i>1$, corresponding to the attitude: nothing but the best. From (15) $\bar{U}(1, r)=r/n$ and $\bar{U}(s, r)=0$ for $s>1$. This last result means that it is no good proposing to anyone of apparent rank other than one, for to do so will result in zero utility. Consequently all we have to do is to find out when it pays to propose to a lady who has apparent rank 1. Equation (16) gives

$$U(1, r) = \max\left\{r/n: \sum_{i=1}^{r+1} U(i, r+1)/(r+1)\right\} \quad \ldots\ldots (17)$$

and

$$U(s, r) = \sum_{i=1}^{r+1} U(i, r+1)/(r+1) \text{ for } s>1 \quad \ldots\ldots (18)$$

Since s does not appear on the right-hand side of (18), $U(s, r)$ must, for $s>1$, be a function only of r, u_r, say. Equations (17) and (18) may then be written

$$U(1, r) = \max\{r/n, u_r\} \quad \ldots\ldots (19)$$

and

$$u_r = \frac{1}{r+1}\{U(1, r+1)+ru_{r+1}\} \quad \ldots\ldots (20)$$

respectively. Now suppose $U(1, r)>\bar{U}(1, r)=r/n$, that is the utility of continuing exceeds that of proposing. It follows from (19) that $U(1, r)=u_r$ and, from (20) with the value of r reduced by one, that $u_{r-1}=u_r$. Therefore since $u_r>r/n$, $u_{r-1}>(r-1)/n$ and from (19), again with the value of r reduced by one, $U(1, r-1)>(r-1)/n$. Hence if it is not worth proposing to a lady who is best out of r, it is not worth proposing to a lady who is best out of $(r-1)$, which result is intuitively obvious. Hence the best strategy must be to propose to a lady who is best out of r, provided r is large enough. We have only to find out how large r must be.

Suppose now that $U(1, r)=\bar{U}(1, r)=r/n$; that is, it is worth proposing to the best lady out of r, and therefore $U(1, r')=r'/n$ for all $r'\geq r$, by what we have just proved. Equation (20) gives

$$u_r = \frac{1}{r+1}\left\{\frac{r+1}{n}+ru_{r+1}\right\}$$

or, if $v_r=u_r/r$,

$$v_r = \frac{1}{nr}+v_{r+1}$$

and the same equation will obtain for all larger r. Adding both sides of all these equations we have

$$v_r = \frac{1}{n}\left\{\frac{1}{r}+\frac{1}{r+1}+\ldots+\frac{1}{n-1}\right\} \quad \ldots\ldots (21)$$

But we know that u_r must be less than r/n since $U(1, r)=r/n$ [equation (19)], so that $v_r<n^{-1}$. Hence

$$\frac{1}{r}+\frac{1}{r+1}+\ldots+\frac{1}{n-1}<1 \qquad \ldots\ldots (22)$$

The argument may be reversed and consequently it is only worth proposing to the apparently best lady out of r if (22) obtains. Let $r=R$ be the least such value; that is,

$$\frac{1}{R}+\frac{1}{R+1}+\ldots+\frac{1}{n-1}<1<\frac{1}{R-1}+\frac{1}{R}+\frac{1}{R+1}+\ldots+\frac{1}{n-1}$$

then from (20) with $r=R-1$

$$u_{R-1}=\frac{1}{R}\left\{\frac{R}{n}+(R-1)u_R\right\}$$

$$=\frac{R-1}{n}\left\{\frac{1}{R-1}+\frac{1}{R}+\ldots+\frac{1}{n-1}\right\}\ldots\ldots (23)$$

from (21). Again from (20) with $r<(R-1)$ we have $u_r=u_{r+1}$, so that u_1, the utility of the best strategy starting at the beginning, is given by (23). If n is large the value of R is given by $\int_R^n dx/x=1$, if the series in (22) is replaced by an integral, and hence $n/R=\mathrm{e}$, the base of natural logarithms. Hence for large n the optimum rule is to inspect until a proportion e^{-1} $(=0{\cdot}368)$ of the ladies have been inspected and then propose to any subsequent lady of apparent rank one; the expected utility is, from (23), only e^{-1}. If, in real life, this process works between 18 and 40 (i.e. for 22 years) one should never propose until age $18+0{\cdot}368\times 22=26$, approximately. Either many people do not pursue an optimum strategy or else they have a different utility function.

To learn something of the effect of a change in utility we consider a second case in which $U_i=n-i$; that is, linear in the rank. It is a simple matter to calculate

$$\bar{\mathrm{U}}(s,r)=n-\left(\frac{n+1}{r+1}\right)s \qquad \ldots\ldots (24)$$

and this value has to be inserted in (16). Now since $\bar{\mathrm{U}}(s,r)$ decreases as s increases and

$$\sum_{i=1}^{r+1}\mathrm{U}(i,r+1)/(r+1)$$

in (16) does not depend on s it is clear from (16) that if it is worth proposing to a lady of apparent rank s out of r, it is certainly worth proposing to one of apparent rank s' out of r, where $s'<s$; this is intuitively obvious. So the optimum strategy can be defined by a function $\mathrm{s}(r)$ with the rule: propose if and only if $s\leqslant \mathrm{s}(r)$; $\mathrm{s}(r)$ is the boundary in the sense used on p. 45. It follows from (16), again because

$$\sum_{i=1}^{r+1}\mathrm{U}(i,r+1)/(r+1)$$

does not depend on s, that $U(s, r)$, for $s > s(r)$, does not depend on s; so let us write it $n - c(r)$, say. From (16) we obtain the two equations

$$U(s, r) = \max\left\{n - \left(\frac{n+1}{r+1}\right)s,\ n - c(r)\right\} \quad \ldots\ (25)$$

and

$$n - c(r-1) = \sum_{i=1}^{r} U(i, r)/r \quad \ldots\ (26)$$

where, in (26), the value of r has been decreased by 1. Since one will propose to a lady of rank $s(r)$ but not of rank $s(r)+1$, (25) gives

$$\left(\frac{n+1}{r+1}\right)[s(r)+1] \geq c(r) \geq \left(\frac{n+1}{r+1}\right)s(r) \quad \ldots\ (27)$$

and (26) gives

$$c(r-1) = \left\{\left(\frac{n+1}{r+1}\right)[1+2+3+\ \ldots\ +s(r)] + [r - s(r)]c(r)\right\}\frac{1}{r} \quad \ldots\ (28)$$

These two results are the most convenient form for calculation purposes. But the calculations are rather tedious and, as emphasised on p. 45, it would be highly desirable to have a result for $s(r)$ alone. Although it is difficult to obtain a differential equation for the boundary in the normal case on p. 45, it is easy to do so here. Suppose n is large and let $x = r/n$; x is then the proportion of ladies so far inspected. Allow n to approach infinity but x remain fixed (so that also $r \to \infty$). Let $C(x) = c(r)$ and $S(x) = s(r)$. Then (27) is approximately

$$C(x) = S(x)/x \quad \ldots\ (29)$$

and (28) is approximately

$$c(r-1) - c(r) = \left\{\frac{1}{x} \cdot \frac{[S(x)]^2}{2} - S(x)C(x)\right\}\frac{1}{xn} \quad \ldots\ (30)$$

(In the last result we have used the fact that the sum of the first n natural numbers is $n(n+1)/2$. But $c(r-1) = C(x - n^{-1})$ so that, from (30),

$$n\{C(x - n^{-1}) - C(x)\} = -\tfrac{1}{2}[C(x)]^2$$

or, as $n \to \infty$,

$$dC/dx = \tfrac{1}{2}[C(x)]^2 \quad \ldots\ (31)$$

The boundary condition is $C(1) = c(n) = \frac{1}{2}n$, and therefore the solution to (31) is

$$C(x) = \{2n^{-1} + \tfrac{1}{2}(1 - x)\}^{-1} \quad \ldots\ (32)$$

and therefore, from (29)

$$S(x) = x\{2n^{-1} + \tfrac{1}{2}(1 - x)\}^{-1} \quad \ldots\ (33)$$

This approximation ignores the discrete form of $s(r)$, but numerical computations suggest that it is quite good. For example, with $n = 100$

and $r=90$, so that $x=0{\cdot}9$, $S(x)=90/7$, or about 13. The exact value is 14, though the change in utility by using 13 instead of 14 is negligible. Notice that the approximation is only valid for fixed x as n gets large. No proposals should be made until $S(x)>1$, that is until $x>1/3+4/3n$. For large n this gives $x>1/3$, agreeing very well with the strategy in the first case where the corresponding value was 0·368. $C(0)=2n/(4+n)$ so that the expected utility at the start is $n-2n/(4+n)=n(n+2)/(n+4)$. This is only accurate to order 1 and to this order is $(n-2)$, so that the expected utility is nearly equal to that of the best lady, namely $(n-1)$. This approximation is, however, suspect. One could also argue that since no proposals are to be made until $x>1/3+4/3n$ the expected utility is constant until this inequality is first satisfied and therefore that $n-C(1/3+4/3n)$ is an equally good approximation to the expected utility at the start. To order 1 this gives $(n-3)$.* Mr B. N. Barnett has suggested alternative approximations to the behaviour of $C(x)$ for small x which appear to be better, but more work remains to be done before a completely satisfactory result can be obtained.

This problem also arises in the game of 'googol' mentioned in the *Scientific American* for February 1960. One player writes n different, but otherwise arbitrary, numbers on n pieces of paper and places them face downwards on a table. His opponent turns them up one at a time and can stop at any stage and say that the one just turned up is the biggest. If correct he wins, otherwise he loses. The first utility function is relevant.

* I am indebted to the Editor for this idea.

REFERENCES

[1] Armitage, P. (1960). *Sequential Medical Trials*. Blackwell, Oxford.

[2] Bellman, R. (1957). *Dynamic Programming*. Oxford University Press.

[3] Box, G. E. P. and Wilson, K. B. (1951). 'On the experimental attainment of optimum conditions', *J. R. Statist. Soc.*, B, **13**, 1.

[4] Chernoff, H. (1960). *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. (To appear.)

[5] Grundy, P. M., Healy, M. J. R. and Rees, D. H. (1956). 'Economic choice of the amount of experimentation', *J. R. Statist. Soc.*, B, **18**, 32.

[6] National Bureau of Standards (1951). *Tables to Facilitate Sequential t-Tests*. Appl. Math. Series 7. Government Printing Office, Washington.

[7] Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

[8] Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.

[9] Wald, A. and Wolfowitz, J. (1948). 'Optimum character of the sequential probability ratio test', *Ann. Math. Statist.*, **19**, 326.

A METHOD OF ALLOWING FOR 'NOT-AT-HOME' BIAS IN SAMPLE SURVEYS

D. J. BARTHOLOMEW
*University College of North Staffordshire**

In this article Dr Bartholomew describes a method for reducing bias in the estimation of a proportion or a mean arising from failure to contact respondents who are not at home. It depends upon an assumption which has been tested using actual survey data.

Introduction

When a survey is based on a random sample of individuals, standard statistical theory enables us to make inferences about the population from which the sample was drawn. In practice this process is frequently invalidated by the failure to obtain a response from every member of the sample. In such cases we have no guarantee that the *working* sample is random and hence the estimates obtained from it may be biased. Moser[1] has distinguished six different sources of bias which occur in surveys conducted by interviewers calling at the homes of respondents. These include: inadequacy of the sampling frame, refusal to provide the required information, and the respondent being out at the time of the call. In the present paper we shall be concerned with bias from the last-named source. In theory it can be eliminated by repeated call-backs but in practice this is seldom possible. Repeated calls increase the cost of the survey, delay its completion, and may have a detrimental effect on the interviewers' morale. It is thus very desirable to have some method of eliminating or reducing 'not-at-home' bias without having to make unlimited calls. The new method which I propose is very simple. It involves eliminating all bias which enters at the first call. I shall advance theoretical and empirical reasons for believing that most if not all bias arises at this stage. It will thus be possible to obtain estimates which are practically unbiased after only two calls.

Before proceeding to discuss this in detail I shall briefly outline other methods which have been devised to achieve the same end. It is common practice among survey practitioners to insist on three, or at least three, calls on each respondent. In an experimental survey covering a large number of topics, Durbin and Stuart[2] have shown that this procedure did not lead to serious bias. Hansen and Hurwitz[3] suggested repeated call-backs on a sub-sample of those not contacted at the first call and have indicated how the size of the sub-sample should be chosen. This method has considerable attractions if the call-back sample is small, but these must be set against the relatively large

* Present address: University College of Wales, Aberystwyth.

standard errors of the sample estimates. It was introduced to avoid the high cost of interviewing respondents who failed to return mailed questionnaires. The basic cause of not-at-home bias is that the probability of finding the respondent at home is correlated with the subject of the enquiry. This led Politz and Simmons[4] to suggest that questions should be inserted in the schedule to enable this probability to be estimated. Each respondent is asked whether he was at home at the time of the interview on each of the preceding five (or six) days. This information can then be used to stratify the sample members according to their chance of being found at home and thus enables unbiased estimates to be obtained. The method was tried out by Durbin and Stuart[2] in their experimental survey but the results were disappointing. However, it must be noted that in their survey calls were made at all times of the day, instead of in the evenings as the authors of the method suggest. Another disadvantage is that the Politz-Simmons method requires a considerable effort of memory on the part of the respondent, which may result in inaccurate reporting. The disadvantages of these methods are avoided in the new approach, which gives very promising results in practice.

The Estimation of a Proportion

Suppose that we wish to estimate the proportion of people in the original sample who possess the attribute A. Let this proportion be denoted by P and the sample size (the total number who could be contacted with re-calling continued indefinitely) by $\mathcal{N}$. We require an estimate of P based on the results obtained at the first two calls; these will be referred to as the 'first call' sample and the 'second call' sample. It must be remembered that P itself will be an estimate of the proportion of A's in the population. Let the number of respondents at home at the first call be $\mathcal{N}_1$, of whom n_1 are found to possess A. Suppose that the corresponding numbers at the second call are $\mathcal{N}_2$ and n_2. The quantities $\mathcal{N}$, $\mathcal{N}_1$, $\mathcal{N}_2$, n_1, and n_2 are known; P is to be estimated.

Let $P_1 = n_1/\mathcal{N}_1$ (the proportion of A's in the first call sample) and let $P_r = (\mathcal{N}P - n_1)/(\mathcal{N} - \mathcal{N}_1)$, i.e. the proportion of A's left in the original sample after the first call has been made. Then clearly

$$P = \frac{\mathcal{N}_1}{\mathcal{N}} P_1 + \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right) P_r \qquad \ldots\ldots (1)$$

After the first call P_1 is fixed, so all that is required is an estimate of P_r. If bias is present only at the first call this will be provided by $\hat{P}_r = n_2/\mathcal{N}_2$. Thus we shall have for our estimate of P

$$\hat{P} = \frac{\mathcal{N}_1}{\mathcal{N}} P_1 + \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right) \hat{P}_r$$

$$= \frac{1}{\mathcal{N}} \left\{ n_1 + \frac{n_2(\mathcal{N} - \mathcal{N}_1)}{\mathcal{N}_2} \right\} \qquad \ldots\ldots (2)$$

This estimate is quite independent of any first call bias. Since I shall argue below that this is the main, if not the sole, source of bias in survey work we may expect (2) to provide an almost unbiased estimate of P after only two calls.

The sampling variance of $\hat{P}$ will be

$$\text{var}\,(\hat{P}) = \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right)^2 \text{var}\,(\hat{P}_r)$$

Now $\hat{P}_r$ is obtained, we assume, by random sampling from a finite population, so the familiar result applies, thus

$$\text{var}\,(\hat{P}) = \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right)^2 \left(\frac{\mathcal{N} - \mathcal{N}_1 - \mathcal{N}_2}{\mathcal{N} - \mathcal{N}_1 - 1}\right) \frac{P_r(1 - P_r)}{\mathcal{N}_2} \quad \ldots . \quad (3)$$

In practice P_r will be unknown and must be replaced by its sample estimate $\hat{P}_r$. If we wish to test hypotheses or make estimates for the whole population we must use [var $(\hat{P})$ + var (P)] where var (P) is the sampling variance of P based on a sample of $\mathcal{N}$ from the population under investigation.

The results may be immediately generalised though this is of limited practical value. Suppose that a maximum of k calls are made and that from the rth call onwards $(r \leqq k)$ sampling is effectively random. The formulae given above still apply if $\mathcal{N}_1$ and n_1 are taken to apply to the combined result of the first $(r-1)$ calls and $\mathcal{N}_2$ and n_2 to the remainder.

Estimation of the Mean

The foregoing analysis can be regarded as a special case of estimating the mean. Suppose that, for our sample of size $\mathcal{N}$, some measurable characteristic takes the values $x_1, x_2, \ldots x_{\mathcal{N}}$. Let

$$\mu = \sum_1^{\mathcal{N}} x_i/\mathcal{N} \qquad \mu_1 = \sum_1^{\mathcal{N}_1} x_i/\mathcal{N}_1 \quad \text{and} \quad \lambda = \sum_{\mathcal{N}_1+1}^{\mathcal{N}} x_i/(\mathcal{N} - \mathcal{N}_1)$$

then the sample mean is

$$\mu = \frac{\mathcal{N}_1}{\mathcal{N}}\mu_1 + \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right)\lambda$$

If $x_1, x_2, \ldots x_{\mathcal{N}_1}$ are the values resulting from the first call and if sampling at the second call is effectively random and yields $x_{\mathcal{N}_1+1}$, $x_{\mathcal{N}_1+2}, \ldots x_{\mathcal{N}_1+\mathcal{N}_2}$ then μ may be estimated by

$$\hat{\mu} = \frac{\mathcal{N}_1}{\mathcal{N}}\mu_1 + \left(1 - \frac{\mathcal{N}_1}{\mathcal{N}}\right)\hat{\lambda} \quad \ldots . \quad (4)$$

where $$\hat{\lambda} = \sum_{\mathcal{N}_1+1}^{\mathcal{N}_1+\mathcal{N}_2} x_i/\mathcal{N}_2$$

The variance of $\hat{\mu}$ is given by

$$\text{var}(\hat{\mu}) = \left(1 - \frac{N_1}{N}\right)^2 \left(\frac{N - N_1 - N_2}{N - N_1 - 1}\right) \frac{\sigma^2}{N_2} \quad \ldots . (5)$$

σ^2 is the variance of $x_{N_1+1}, x_{N_1+2}, \ldots x_N$ and would be estimated from the data collected at the second call. If x can take only the values 0 and 1, then (4) and (5) reduce to the corresponding results for a proportion.

The Basic Assumption

We must now discuss the grounds for believing that the second call sample will be a random sample of all respondents not found at the first call. If the interviewer was fully informed of the times each respondent was normally at home she would be able to time her calls so that each member of the sample had approximately the same chance of being contacted. (For simplicity we shall refer to the interviewer as 'she' and the respondent as 'he'.) In this case we might expect to obtain a random sample at the first call. In practice this kind of information is usually lacking, so that the first call sample will be heavily biased in favour of those who spend most of their time at home. Before making the second call, however, the interviewer will have had an opportunity to obtain some information about the habits of the respondent. This information can be acquired by the interviewer from neighbours or other members of the household. It may even be possible to make an appointment for the second call. Thus whatever bias is present at the first call should be considerably reduced, if not eliminated, at the second call.

Since the procedure outlined above is that normally followed in survey work, it should be possible to test our assumption empirically. This has been done for three sets of data and the results are discussed below.

The first set of data was obtained from a survey[5] concerned with local government affairs carried out in Newcastle-under-Lyme by student interviewers. Random samples of electors were drawn from each of four wards in the borough and rural district. If the respondent was out at the first call, the interviewers were instructed to find out from neighbours or other members of the household when he was likely to be at home before calling again.

To test our assumption we need to know what result would have been obtained if recalling had been continued until every member of the sample had been accounted for. For this reason it is convenient to use the sex of the respondent as the attribute under investigation. This could always be determined from the Christian names given on the electoral register, and consequently was known for the whole sample. It has the additional advantage of being related to the chance of being found at home at the first call; hence it leads to a not-at-home bias in the incomplete sample. Table I*a* provides the information required for testing the randomness of sampling at the first and second calls.

TABLE I*a*

Comparison of percentages of men obtained at the first and second calls with those expected on the hypothesis of random sampling (Local Government Survey)

Ward	A	B	C	D
Percentage of Men in First Call Sample	42·3 (52) *4·53*	32·3 (62) *4·61*	31·1 (61) *4·79*	35·4 (48) *5·75*
Percentage of Men in Whole Sample ($100P$)	52·2 (90)	48·5 (130)	49·6 (137)	52·3 (130)
Percentage of Men in Second Call Sample	59·3 (27) *4·98*	61·4 (44) *4·35*	66·7 (51) *3·87*	63·8 (47) *4·65*
Percentage of Men left after First Call ($100P_r$)	65·7 (38)	63·2 (68)	64·5 (76)	62·2 (82)

Each percentage is accompanied by the total frequency (in brackets) on which it is based. The figures in italics in the first and third rows of the table are the standard errors of the percentages calculated on the assumption that at each stage every person has an equal chance of being at home.

TABLE I*b*

Comparison of percentage estimates

Ward	A	B	C	D
$100P$ (Percentage of Men in Whole Sample)	52·2	48·5	49·6	52·3
$100\hat{P}$ (by Method described above)	49·5 *2·15*	47·5 *2·28*	50·8 *2·10*	52·3 *2·89*
$100P_2$ (Percentage of Men at First Two Calls) $\left(=\frac{n_1+n_2}{N_1+N_2}\times 100\right)$	48·1 *1·98*	44·3 2·08	47·3 *2·02*	49·5 *2·67*

In every case the percentage of men in the first call sample is less than $100P$ by more than twice its standard error. We thus conclude that there is a clear bias at this stage. Turning now to the data on the second call, we find a close agreement between the observed and expected percentages. In only one case is the difference more than one standard error (1·28 S.E.). There is thus no reason to doubt the validity of the assumption and on this evidence we should expect $\hat{P}$ to provide a good estimate. This is shown in Table I*b* which compares a number of estimates. In every case $\hat{P}$ is very much closer to the true P than is P_2, which is calculated as if the combined result of the first two calls provided a random sample. The figures in italics are the appropriate standard errors. In comparing these it must be remembered that those for P_2 are based on the assumption that the first two calls are random, in which case there would be no need for $\hat{P}$.

Data collected by Edwards,[6] using professional interviewers, provides a more stringent test because the survey involved a very much larger sample. The results are set out in Tables II*a* and II*b*. It is clear once

again that although there was a substantial bias at the first call, this was not so at the second.

TABLE II*a*

Comparison of percentages of men obtained at the first and second calls with those expected on the hypothesis of random sampling (Edwards's data)

Percentage of Men in the First Call Sample	29·5 (1243) *1·091*
Percentage of Men in the Whole Sample	43·6 (3117)
Percentage of Men in the Second Call Sample	51·9 (1146) *0·919*
Percentage of Men Left after the First Call	53·0 (1874)

TABLE II*b*

Comparison of percentage estimates

	Estimate	Standard Error
$100P$	43·6	—
$100\hat{P}$ (after two calls)	43·0	0·55
$100P_3\left(\frac{n_1+n_2+n_3}{N_1+N_2+N_3}\times 100\right)$	42·6	—
$100P_2\left(\frac{n_1+n_2}{N_1+N_2}\times 100\right)$	40·3	0·49

It is interesting to notice that $\hat{P}$ calculated after two calls is closer to P than P_3, which required three calls.

Durbin and Stuart[7] classified the results of the second call in an experimental survey according to whether or not an appointment had been made at the first call. A summary of their results is given in Table III.

TABLE III

Percentage of successes at the second call in an experimental survey; total frequencies are given in parentheses (Durbin and Stuart's data)

Appointment Made		No Appointment	
Men	Women	Men	Women
72·5	69·0	34·6	44·3
(258)	(200)	(188)	(212)

It will be seen that the sex bias, which was present at the first call, was eliminated for those second calls for which an appointment had been made. In the case of no appointment, the difference is significant at the 10% level. Although this is not conclusive, the direction of the difference suggests that some bias is still present. It is not clear whether any information at all was obtained by the interviewers about those for whom no appointment could be made. In the absence of knowledge about the habits of these respondents we should not expect the second call sample to be unbiased. Durbin and Stuart[2] have given further

data which relate not only to sex but to a large number of other attributes. Unfortunately for our purposes the second and third call results have been combined and the original records are no longer available. However, using the generalisation indicated on page 54 with $k=3$, $r=2$, it is possible to compare $\hat{P}$ with P_3. This comparison is of limited value since the bias of P_3 in almost all cases was very small. In no case was $\hat{P}$ significantly worse than P_3 and in many cases it was better.

The evidence which we have examined suggests that the new method is likely to offer a worthwhile improvement over crude estimates based on the same number of calls, especially if appointments can be made when the first call is unsuccessful. It would be very desirable to carry out similar analyses for attributes other than sex. However, we feel that the elimination of first-call bias is sufficiently valuable in itself to justify the use of the new method.

Discussion

The three methods of eliminating 'not-at-home' bias outlined in the introduction are subject to various drawbacks. The first requires at least three calls, the second is liable to give high sampling variances for estimates, and the third will be subject to error if the respondent fails to remember accurately. For the new method, two calls are sufficient and no effort of memory on the part of the respondent is required. The sampling variance of $\hat{\mu}$ and $\hat{P}$ depends on the relative values of N, N_1, and N_2. It is easy to construct examples with a very high variance relative to that of P_2. However, Tables I and II show that for the values of N, N_1, and N_2 used in our examples—which are typical of those met with in general practice—the increase over the random sampling variance of P_2 is very small.

The survey practitioner must therefore decide whether the error which may result from using $\hat{P}$ after the second call is likely to be large enough to justify the time and cost of three or more calls. This cannot be accurately assessed in isolation from the practical problem but our experience suggests that this error is unlikely to be serious.

The method given here will not be applicable unless information can be obtained or appointments made at the first call. Thus it would be quite unsuitable for postal questionnaires. Some progress has been made on the general problem of eliminating bias under wider assumptions. For example, we may suppose that an interviewer makes calls at random, disregarding any information which she may have about the habits of the respondents. The chance of finding any given person at home will then be constant at each call. Progress on these lines is hampered by lack of suitable data to test the realism of the theory.

REFERENCES

[1] Moser, C. A. (1958). *Survey Methods in Social Investigation.* Heinemann, London.
[2] Durbin, J. and Stuart, A. (1954). 'Callbacks and clustering in sample surveys; an experimental study', *J. R. Statist. Soc.*, A, **117**, 387.
[3] Hansen, M. H. and Hurwitz, W. N. (1946). 'The problem of non-response in sample surveys', *J. Amer. Statist. Assoc.*, **41**, 517.
[4] Politz, A. and Simmons, W. (1949). 'An attempt to get the "not-at-homes" into the sample without call-backs', *J. Amer. Statist. Assoc.*, **44**, 9.
[5] Bealey, F. W. and Bartholomew, D. J. (1961). To appear in *British Journal of Sociology.*
[6] Edwards, F. (1953). 'Aspects of random sampling for a commercial survey', *The Incorporated Statistician*, **4**, 9.
[7] Durbin, J. and Stuart, A. (1951). 'Differences in response rates of experienced and inexperienced interviewers', *J. R. Statist. Soc.*, **114**, 163.

NOTES AND COMMENTS

Readers of this journal may well have wondered why the actual date of publication lags so far behind that printed on the cover. Originally some disorganisation of the schedule occurred as a result of the 1959 printing dispute, and such disorganisations often have far-reaching effects, but the main cause of the present delays is the shortage of suitable papers submitted in 1960. It may seem strange in these days, when so many journals are short of publication space, to be appealing for articles, but it is a sad fact that the supply has not been as plentiful as we should like.

The situation has recently shown some improvement with respect to the number of articles submitted, but sometimes authors appear to have an imperfect understanding of the type of article we require; the tendency is for the material offered to be too theoretical. The following suggestions give some indication of the types of article that are welcome:

(*a*) case-histories of interesting practical applications of statistics;
(*b*) expositions of new methods or of new applications of established methods;
(*c*) articles of general interest to all readers, such as surveys of the applications of statistics in broad fields, general descriptions of the statistical set-up in institutions and firms, etc.

The fields of application covered in this journal are quite wide: in 1960 they included such diverse subjects as geology, psychiatry, spectrochemical analysis, and the economics of new roads. We do not lack variety; we *do* lack sufficient quantity. We urge all statisticians, particularly those working in this country, to take up their pens; a simple description of some piece of work, which otherwise would be known to only a limited few, might well be helpful to statisticians in other parts of the world.

MEETINGS OF SECTIONS AND LOCAL GROUPS OF THE ROYAL STATISTICAL SOCIETY

GENERAL APPLICATIONS SECTION

11th October 1960. Professional Earnings: W. F. F. Kemsley

11th November 1960. Some Aspects of the Radcliffe Committee's Report: H. B. Rose

13th December 1960. Operational Research and Statistics: A Comparison and a Criticism: B. H. P. Rivett

INDUSTRIAL APPLICATIONS SECTION

4th November 1960. Estimation of Extremes: E. H. Lloyd. Aspects of Production Planning and Scheduling: W. N. Jessop

MEDICAL SECTION

27th September 1960. Multiple Regression Analysis applied to Records of Sepsis following Operation: O. M. Lidwell

25th October 1960. Closed Sequential Designs in Clinical Research: C. C. Spicer

29th November 1960. Biometrical Problems in Studies of Human Growth: J. M. Tanner

BIRMINGHAM AND DISTRICT GROUP

4th October 1960. Estimation of Extremes: E. H. Lloyd

20th October 1960. Some Aspects of Educational Statistics: D. M. Lee

9th November 1960. Statistics and Work Measurement: P. Fisken (Joint Meeting with the Institute for Work Study)

21st November 1960. Public Opinion Polls: C. A. Moser

6th December 1960. The Analysis of Censored Data: J. Taylor

BRISTOL GROUP

20th October 1960. Practical Experimentation in the Teaching of Statistics: Hilda M. Davies

17th November 1960. Use of Statistics in Management: P. J. Stanley

15th December 1960. Statistical Applications in the Chemical Industry: P. J. Harrison

EAST MIDLANDS GROUP

20th October 1960. Interfirm Comparison for Management: F. Conway

10th November 1960. Statistical Models in Operational Research: L. Taylor-Harrington

8th December 1960. The Impact of Computers on Statistics: A. Muir

EDINBURGH GROUP

5th October 1960. Simulation, Monte Carlo, and Industrial Problems: W. N. Jessop

21st November 1960. Statistical Methods as Aids to Business Management: T. Hunter-Thomson

GLASGOW GROUP

20th October 1960. Aspects of Statistical Quality Control: L. Irvine
23rd November 1960. Some Problems of Sampling for Large-scale Enquiries: P. J. Lane
13th December 1960. Symposium on Industrial and Commercial Applications of Computers. Computers—an Introduction: J. G. Fraser; Computers in Sales Control: C. Hossell; Computers in Industry and Design: N. Leby

MERSEYSIDE GROUP

11th October 1960. The Analysis of Censored Data: J. Taylor
8th November 1960. Economics of Routine Testing in the Chemical Industry: O. L. Davies
13th December 1960. Evolutionary Operation in Practice: G. A. Coutie

NORTH-EASTERN GROUP

12th October 1960. Unusual Frequency Distributions: D. J. Newell
9th November 1960. The Binomial Hypothesis in Mendelian Inheritance: D. A. Evans
14th December 1960. The Basis, Purpose, and Applications of Information Theory: C. Cherry

SHEFFIELD GROUP

27th October 1960. Measuring Economies of Scale: C. B. Winsten
30th November 1960. New Uses of Computers in Statistics: G. A. Barnard

SOUTH WALES GROUP

6th October 1960. Statistical Methods in Auditing: P. N. Wallis
1st November 1960. Accident Rates on Rural Trunk Roads: J. R. Scott
7th December 1960. Analysis of Censored Data: J. Taylor

BOOK REVIEWS

Emanuel Parzen: *Modern Probability Theory and Its Applications*. New York and London: John Wiley and Sons, 1960. Pp. xv+464. 86s.

It is as well to state at the outset that this book is not aimed primarily at the type of reader for which this journal is supposed to cater. It is written as a textbook for a course in probability that can be adapted to the needs of students with diverse interests and backgrounds. I think it is possible that a student working under supervision would find it quite useful; it contains a good deal of information and is well supplied with exercises of various kinds, some of them being 'theoretical exercises' in which the student is made to do some of the theoretical work himself. I should be less happy in recommending this book to a private student, as I feel that the arguments sometimes lack cohesion at vital points.

The beginning seems rather shaky; is it not asking for trouble to try to define a random event on the second page? We read: 'A random event is one whose relative frequency of occurrence, in a very long sequence of observations of randomly selected situations in which the event may occur, approaches a stable limit value as the number of observations is increased to infinity. . . .' Apart from the circularity inherent in this definition, it does not properly distinguish random from systematically occurring events.

Of course this type of circularity bedevils all attempts to define probability; the present relatively sophisticated approach via sample descriptive spaces still postulates the ability to recognise situations in which it may be assumed that all descriptions are 'equally likely' in order to evaluate probabilities. I am not worried by this. What does worry me is that, possibly because of an endeavour to preserve mathematical rigour, the author tends to confuse the reader. Thus on page 25 he implies that there are cases in which it may reasonably be assumed that all descriptions are equally likely; on page 26 he implies that the correctness of such a mathematical model can never be decided *a priori* but only after performing experiments.

On page 69 the equation is inappropriate to the physical situation described, because the mathematical model is incorrect. I confess to great difficulty in attributing a useful physical meaning to drawing an unordered sample of size n with replacement; surely the idea of order is implicit in drawing with replacement. It is true that after the drawing the order can then be ignored, but if that is done the sample descriptions are no longer equally likely. A similar situation involving drawing an unordered sample with replacement is compared to Bose-Einstein statistics and here again I feel that the comparison is likely to mislead.

However, despite these blemishes there is much useful material for the student. What is there for the statistician? There is quite a lot on probability laws of various kinds, means and variances, generating functions, the law of large numbers, the central limit theorem, and similar topics. Various applications are also discussed, such as sampling inspection, telephone waiting-times, haemocytometer counts, matching problems, and the duration of games.

Although significance tests and the idea of some kind of confidence interval are implicitly touched on in some parts of the book, no explicit treatment of them is given. It is perhaps unfair to criticise the author for not doing what he had no intention of doing; there is a limit to the topics that can be included in such a book as this. Yet the fact must be faced that a book on probability and its applications which omits such important topics as confidence limits and fiducial limits leaves the statistician's appetite incompletely satisfied. D. G. BEECH

Norbert L. Enrick: *Quality Control (4th Edition)*. New York: The Industrial Press; Brighton: Machinery Publishing Co. Ltd., 1960. Pp. 216. $5.50.

This book seeks to be the practical man's guide to Quality Control, with emphasis on 'simplified presentation from a practical viewpoint'; such a book is badly needed, but this one does not meet the need. It has a somewhat old-fashioned—even Georgian—flavour; little in it was new when the first edition came out in 1948. This is not, in itself, a worthwhile criticism to make, except that this old-time outlook is associated with a lack of insistence upon a very important point. Quality Control, in all its ramifications, is essentially a feedback device which depends, in the last resort, upon someone taking appropriate action on the plant whenever the warning signals are displayed (and, contrariwise, refusing to take action until there is good reason to do so). Unless this point is taken, Quality Control merely provides ingenious recording devices for collecting samples of pointless information. A book on Quality Control can get by with very little statistics, and it may even be forgiven some statistical errors, but it must always stress the fact that a system can only prove its worth insofar as proper technical action flows from its operation. Unhappily, this book, although containing many admirable references to matters of organisation of quality control systems, fails to deal with this central theme, and this makes it, in my opinion, not the book to put in the hand of the practical man just about to meet Quality Control for the first time. W. A. PRIDMORE

A CRITICAL EXAMINATION OF SIMON'S MODEL OF CERTAIN DISTRIBUTION FUNCTIONS IN LINGUISTICS

GUSTAV HERDAN
Lecturer in Statistics, University of Bristol

In this article Dr Herdan considers some of the pitfalls which lie in the paths of those who construct mathematical models, with particular reference to the study of linguistics. He then shows how some of the properties of word frequency distributions can be explained.

For mathematical models to be of real value it is necessary that (1) the relationship between events of which the mathematical structure is to be a model should be what the mathematician believes it to be; (2) that the assumptions needed for constructing the model should be sensible, i.e. in accordance with how the operations in question take place; and (3) that the formulae derived in this way should fit the observed facts. None of these requirements must be neglected if the model is to fulfil its purpose. It is now a sad fact that model construction in mathematical linguistics seems dogged by the neglect of one or other of these requirements, especially the first, which cannot but have in its wake the neglect of the other two.

Since the mathematics of these model structures are, as a rule, beyond the non-mathematical linguist, such models are often taken on trust, even by linguists with some experience of statistical procedures. If that trust is misplaced because the model is not really what it purports to be, then the model, instead of fulfilling a useful purpose and encouraging practical numerical work on language data, will prove sterile. This has happened with the Zipf-Mandelbrot law and, more recently, with Simon's model of certain skew distribution functions in linguistics.[1]

Characteristics of Simon's Distribution

Simon starts by describing the characteristics of the skew distributions for which he wishes to construct a 'model'. He ascribes to them three main characteristics:

(a) They are J-shaped in reverse or, at least, highly skew, with very long upper tails. The tails can generally be closely fitted by an empirical function of the form

$$f(i) = (a/i^k)^{b^i} \qquad \ldots . (1)$$

where $f(i)$ is the number of words occurring i times and a, b, and k are constants, b being usually so close to unity that the final factor has no significant effect on $f(i)$, except for very large values of i. Thus,

according to Simon, the number of words that occur exactly i times in James Joyce's *Ulysses* can be fitted by a formula such as a/i^k.

(b) The exponent k is greater than 1, and in the case of word frequencies, is very close to 2; in this case, formula (1) becomes, with b close to unity,

$$\mathrm{f}(i) = a/i^2$$

and

$$i^2\mathrm{f}(i) = a \qquad \ldots . (2)$$

In words: the product of the squared frequency of occurrence and the number of vocabulary items with such frequency is a constant for all values of the frequency.

(c) In the case of word frequencies, the function (1) describes the distribution not merely in the tail, but also for very small values of i. In this case, the ratio of *hapax legomena** to total vocabulary, $\mathrm{f}(1)/V$, where $V = \sum_1^n \mathrm{f}(i)$, is generally in the neighbourhood of $\frac{1}{2}$, and the rate of *dislegomena* to *hapax legomena*, $\mathrm{f}(2)/\mathrm{f}(1)$, is generally in the neighbourhood of $\frac{1}{3}$.

These are, according to Simon, the characteristics of word frequency distributions.

Remembering requirement (1) for the construction of satisfactory mathematical models, we ask: 'does this picture of word counts correspond to the facts?' As will be now shown in detail for criteria (b) and (c), it does not.

I have not aimed at 'random illustration' from the extensive literature in the field, but the following selection of instances from four different languages: Biblical Greek, Latin, English, and Russian, is as far from any conceivable bias as I can make it.

In order to ascertain the extent to which Simon's characteristic (b) and, in particular, his dictum that in the case of word frequencies k is close to 2, correspond to the facts, we shall now apply formula (2) to four different word counts (see Table I).

It is evident from Table I that Simon's characterisation of word count distributions under (b) does not correspond to the facts. The products $i^2\mathrm{f}(i)$ are far from being constant for a given word count; in fact, the quantity increases systematically in the first terms of the series and the tail values are greatly in excess of the other terms. In addition the tail values are at variance with Simon's statement that for word frequencies function (1) also fits the extended tails.

Regarding criterion (c), it may be seen from Table II that although the values given by Simon for $\mathrm{f}(1)/V$ and $\mathrm{f}(2)/\mathrm{f}(1)$, viz. $\frac{1}{2}$ and $\frac{1}{3}$ respectively, are not closely reflected in the empirical results, yet, by and large (with the exception of the New Testament) $\mathrm{f}(1)/V$ appears to lie between 0·4 and 0·5, and $\mathrm{f}(2)/\mathrm{f}(1)$ between 0·3 and 0·4. I shall return to this point later.

* *Hapax legomena* are words used only once in a sample, in a text or whatever the universe of discourse under consideration; *dislegomena* are words used twice, *trislegomena* are words used three times, and so on.

TABLE I

In this table, which refers to criterion (b), $i^2f(i)$ is calculated for the first ten and the last five items of certain word frequency distributions

i	*New Testament* (Greek)[2]	*Imitatio*[3]	Macaulay's *Essay on Bacon*[3]	Pushkin[4,5]: *The Captain's Daughter*
1	$1\times1934=1934$	$1\times520=520$	$1\times990=990$	$1\times2384=2384$
2	$4\times842=3368$	$4\times174=696$	$4\times367=1468$	$4\times847=3388$
3	$9\times470=4230$	$9\times111=999$	$9\times173=1557$	$9\times433=3897$
4	$16\times322=5152$	$16\times70=1120$	$16\times112=1792$	$16\times238=3808$
5	$25\times221=5525$	$25\times37=925$	$25\times72=1800$	$25\times146=3650$
6	$36\times173=6228$	$36\times33=1188$	$36\times47=1692$	$36\times114=4104$
7	$49\times123=6027$	$49\times20=980$	$49\times41=2009$	$49\times82=4018$
8	$64\times110=7040$	$64\times28=1792$	$64\times31=1984$	$64\times79=5056$
9	$81\times107=8667$	$81\times11=891$	$81\times34=2754$	$81\times41=3321$
10	$100\times79=7900$	$100\times14=1400$	$100\times17=1700$	$100\times39=3900$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Last	$2713^2\times1$	$172^2\times1$	$65^2\times1=4225$	$479^2\times1$
	$2771^2\times1$	$196^2\times1$	$76^2\times1=5776$	$528^2\times1$
five	$5534^2\times1$	$210^2\times1$	$81^2\times1=6561$	$724^2\times1$
	$8947^2\times1$	$242^2\times1$	$89^2\times1=7921$	$777^2\times1$
items	$19734^2\times1$	$418^2\times1$	$255^2\times1=65025$	$1160^2\times1$

TABLE II

This table, which refers to criterion (c), gives the ratios $f(1)/V$ and $f(2)/f(1)$ for the four works used in Table I

	New Testament (Greek)	*Imitatio*	Macaulay's *Essay on Bacon*	Pushkin: *The Captain's Daughter*
$f(1)/V$	1934/5436=0·36	520/1168 = 0·45	990/2048 = 0·48	2384/4783=0·50
$f(2)/f(1)$	842/1934=0·44	174/520 = 0·33	367/990 = 0·37	847/2384=0·36

Testing of Simon's Distribution Equations

On the basis of his assumptions about word frequency distributions, Simon arrives at the expression

$$f(i) = A\,.\,B(i, \rho+1) \qquad \ldots\ldots (3)$$

where A and ρ are constants and $B(i, \rho+1)$ is the Beta function of $i, \rho+1$:

$$B(i, \rho+1) = \int_0^1 \lambda^{i-1}(1-\lambda)^{\rho}d\lambda = \frac{\Gamma(i)\,\Gamma(\rho+1)}{\Gamma(i+\rho+1)} \qquad \ldots\ldots (4)$$

$$(0<i;\ 0<\rho<\infty)$$

Since for large i, and for any constant c

$$\frac{\Gamma(i)}{\Gamma(i+c)} \simeq i^{-c} \qquad \ldots\ldots (5)$$

we have from (4), as $i \to \infty$

$$f(i) \simeq A\{\Gamma(\rho+1)\}i^{-(\rho+1)} \qquad \ldots\ldots (6)$$

Therefore, Simon says, the distribution (3) approximates the distri-

bution (1), through the range in which b of (1) is close to 1, which he regards as empirically true for word distributions. Further, if ρ is positive, k will be greater than 1, as required by criterion (b), and if ρ is equal to 1, k will be equal to 2. In the latter case we shall have, from equation (3)

$$f(i) = \frac{1}{i(i+1)} \text{ when } \sum_{i=1}^{\infty} f(i) = 1 \quad \ldots\ldots (7)$$

so that since $f(1) = \frac{1}{2}$ and $f(2) = \frac{1}{6}$, $f(2)/f(1) = \frac{1}{3}$, and $f(1)/\Sigma f(i) = \frac{1}{2}$, as required by (c).

We shall now test what may be regarded as the most important result of Simon's derivation, indeed as the embodiment of his 'model'. From formula (7) we have

$$i^2 f(i) + i f(i) = 1 \quad \ldots\ldots (8)$$

Since the sum on the left of formula (8) is identically the same for every i, summing over all i will give on the right n, if n is the number of word classes according to occurrence frequency i.

In order to express (8) in terms of the absolute frequencies of words with particular occurrence i, we multiply by the total of vocabulary items in the sample, $\Sigma f(i) = V$, and obtain

$$i^2 f(i) + i f(i) = V \quad \ldots\ldots (9)$$

where $f(i)$ now stands for the *number* of words with occurrence frequency i.

Summing over all i, we get

$$\Sigma i^2 f(i) + \Sigma i f(i) = nV \quad \ldots\ldots (10)$$

But the terms on the left occur in the numerator in the formula for the quantity K defined by Yule:[3]

$$K = \frac{\Sigma i^2 f(i) - \Sigma i f(i)}{[\Sigma i f(i)]^2} \quad \ldots\ldots (11)$$

and are, therefore, equal to $K[\Sigma i f(i)]^2 + 2\Sigma i f(i)$ or $KS_1^2 + 2S_1$, if by S_1 we denote the first moment of the word count distribution. We can

TABLE III

Test of Simon's Equation

	K	nV	$KS_1^2 + 2S_1$
Nouns in *Imitatio*	0·008 42	68 × 1168 = 79 424	586 068·3
Nouns in Macaulay's *Essay on Bacon*	0·002 72	50 × 2048 = 102 400	701 833·9
Nouns in *St. John's Gospel*, A.V.	0·016 15	35 × 353 = 12 355	86 112·6
All words in Pushkin: *The Captain's Daughter*	0·006 03	84 × 4783 = 401 772	4 986 270·4

therefore test nV against the value which is equal to the left-hand side of Simon's equation (10). If Simon's equation (8) were true for the whole range of i, then nV must be equal to $KS_1^2 + 2S_1$ which is, in fact, calculated from the whole range of i. (See Table III.)

The discrepancy between theory (Simon's formula) and observation is such as to invalidate Simon's claim that his model really fits word frequency distributions in the whole range of the variable.

In order to see whether Simon's formula, in spite of being unsuitable for describing the whole word count distributions, may yet have some restricted validity for certain values of i, we list in Table IV the quantities calculated by formula (9) in the range $i=1$ to $i=10$. These values may be regarded as theoretical estimates of V derived from the quantities in Table I. Each value listed there in lines 1-10 has been increased by $if(i)$ to give Table IV. The order of magnitude of the five last values in each series is such that the increase can have no appreciable effect and they are, therefore, not listed. Inspection of Table IV shows that the theoretical series of V-values can hardly be regarded as constant, nor does the sum of the $i^2f(i)$ and $if(i)$ for every i equal the number of vocabulary items, as suggested by Simon's formula. However, for a small range of i, say between $i=1$ and $i=5$, there is some degree of stability in the theoretical values of V, which here also appear to be more or less close to the observed vocabulary size.

TABLE IV

This lists the theoretical values of V calculated by formula (9)

i	*New Testament* (Greek)	*Imitatio*	Macaulay's *Essay on Bacon*	Pushkin: *The Captain's Daughter*
1	3868	1040	1980	4768
2	5052	1044	2202	5082
3	5640	1332	2076	5196
4	6440	1400	2240	4760
5	6630	1110	2160	4380
6	7266	1386	1974	4788
7	6888	1120	2296	4592
8	7920	2016	2232	5688
9	9630	990	3060	3690
10	8690	1540	1870	4290

This makes it quite clear where the fault of Simon's model lies. It was conceived on the basis of a superficial picture of what word counts are like by fixing the attention mostly upon the small values of i.

The fact that the ratios $f(2)/f(1)$ and $f(1)/V$ are not too far from the theoretical values of $\frac{1}{2}$ and $\frac{1}{3}$ as expected by Simon's formula, and the fact that for some of the word counts (e.g. from Pushkin's novel) Simon's formula gives a fairly constant value of V for a small range of i, which value is close to the observed number of vocabulary items, seems to indicate that Simon's formula has some degree of validity for the start of the curve from $i=1$ to, say, $i=5$ or thereabout, but not, as it is claimed, for the whole curve and for the extended tail. This is typical of many attempts at fitting word count distributions by a compara-

tively simple formula: each such formula may be moderately suitable for a particular comparatively small range of the variable, but fails to describe the curve along the whole range of the occurrence of its variable.

From all this the conclusion may be drawn that the whole of the extended distribution which, moreover, changes its shape with time or text length—a fact ignored by Simon—is evidently not a homogenous curve obeying one simple function, but the result of the superimposition of a number of functions.

Simon's Mathematical Model

In order to set up a mathematical model, the mathematician has to make certain assumptions which, if not obvious, must at least be sensible for the phenomenon in question. Since, as we have seen earlier, Simon's ideas about word distributions do not really correspond to the facts, we can hardly expect that the particular assumptions he is going to use for setting up his model will be appropriate to the phenomenon in question. That this is so will now be shown in detail.

Simon makes two assumptions for the construction of his stochastic model:[1]

Assumption 1. The probability that the $(r+1)$th word is a word that has already appeared exactly i times is proportional to $if(i, r)$—that is to the total number of occurrences of all the words that have appeared exactly i times.

Assumption 2. There is a constant probability, α, that the $(r+1)$th word be a new word—a word that has not occurred in the first r words.

Assumptions 1 and 2 describe a stochastic process in which the probability that a particular word will be the next one written depends on what words have been written previously.

Simon's assumptions clearly presuppose not only a continuous text, but a piece from the beginning up to a certain point. Otherwise, it would make no sense to speak of the 'probability of the $(r+1)$th word having appeared exactly i times', and of the 'constant probability α that the $(r+1)$th word be a new word—a word that has not occurred in the *first* r words'.

However, the word frequency distributions which Simon mentions —and many others which he did not consider—are not from continuous text samples; for instance, the word frequency lists for English by E. L. Thorndike and I. Lorge[6] and Eldridge[7] and for French by Van der Beke[8] are composite distributions of samples from quite different texts, and for such distributions Simon's assumptions make no sense, since we cannot speak of the '$(r+1)$th word' in the whole distribution if for each particular text sample that word is a different one. Moreover, even when the sample is from one text only, the method of stratified sampling which is here generally used in order to spread the sample

evenly over the whole text—for which purpose it is culled from a number of disconnected pages—implies that Simon's assumptions do not fit. Moreover, the sample may even be selective in the sense that only nouns are considered (Yule) or verbs or other grammatical categories.

A New Model based upon Vocabulary Connectivity

Now it is a remarkable fact that neither the compounded mass of the word frequency distribution, nor its selectivity as regards grammatical categories, has any disturbing influence upon the characteristic properties of such distributions. More precisely, we may compound different word frequency distributions from the same language at the same time of its development into one overall distribution, or we may split an overall frequency distribution characteristic of a given language into parts, i.e. into partial word frequency distributions, without altering the mathematical characteristics of the distribution function.

There is only one distribution function which has this property, and this fact can be used for identifying the mathematical structure of word frequency distributions. As Feller[9] has shown, it is the compound Poisson distribution which has this property. The compound Poisson distribution:

$$\{h_j\} = e^{-\lambda t} \sum \frac{(\lambda t)^n}{n!} \{f_j\}^{n*} \quad \ldots . (12)$$

where $n*$ denotes the n-fold convolution of f_j, has the generating function:

$$h(s; t) = e^{-\lambda t + \lambda t f(s)} \quad \ldots . (13)$$

which has the remarkable property that

$$h(s; t_1 + t_2) = h(s; t_1) h(s; t_2) \quad \ldots . (14)$$

In an intuitive way we may describe this as follows. With each period of duration t there is associated a random variable with generating function $h(s; t)$ which we call the contribution of that period. The contributions of two non-overlapping periods are independent, which means that a partitioning $t = t_1 + t_2$ of a period into two parts induces a decomposition $X(t) = X(t_1) + X(t_2)$ of the total variable into a sum of two independent variables. This assertion is contained in the following theorem (which is a special case of an important general theorem by P. Lévy concerning arbitrary probability distributions):

> If $\{h\}$ is infinitely divisible, its generating function can be written in the form (13), and it can be shown that *only the compound Poisson* distribution has this property.

Illustrations. In order to demonstrate the divisibility of word frequency distributions with retention of the essential characteristics of the total, we turn our attention to the characteristic ratios $2f(1)/V$ and $6f(2)/V$ for such distributions (Table V).

TABLE V

Divisibility of word frequency distribution for New Testament

	New Testament	*Matthew*	*Mark*	*Luke*	*John*	*Acts*	*Paul*	*Hebrews*	*Revelation*
$2f(1)/V$	$\frac{3868}{5436}$	$\frac{1340}{1691}$	$\frac{1268}{1345}$	$\frac{1942}{2055}$	$\frac{751}{1101}$	$\frac{1886}{2038}$	$\frac{2280}{2648}$	$\frac{1150}{1038}$	$\frac{620}{916}$
	=0·712	=0·792	=0·942	=0·945	=0·681	=0·925	=0·861	=1·108	=0·676
$6f(2)/V$	$\frac{5346}{5436}$	$\frac{1656}{1691}$	$\frac{1314}{1345}$	$\frac{2112}{2055}$	$\frac{718}{1101}$	$\frac{2010}{2038}$	$\frac{2632}{2648}$	$\frac{996}{1038}$	$\frac{990}{916}$
	=0·983	=0·979	=0·976	=1·027	=0·652	=0·986	=0·995	=0·960	=1·080

For the whole *New Testament*, it is advisable to calculate the ratio $6f(2)/V$ not only from the *dislegomena*, but also from the *trislegomena*, according to formula (9), and take the average. We then have $6f(2) = 5052$ and $12f(3) = 5640$, with a mean of 5346, which we then use to recalculate $6f(2)/V$.

We see that the characteristic ratios are roughly the same for the whole *New Testament* and for any of the eight parts considered. If we continue the process of subdivision, for instance, for Paul's Letters, we have Table VI.

TABLE VI

Subdivision of word frequency distribution for Paul's Letters

	Romans	*Corinthians*
$2f(1)/V$	1158/1068 = 1·084	878/967 = 0·908
$6f(2)/V$	942/1068 = 0·882	1044/967 = 1·078

Since we thus find the word frequency distribution function characterised by the property of (*a*) combinability and (*b*) divisibility without altering the essential mathematical characteristics of the distribution function, and since it is only the compound Poisson distribution which has this property, we must conclude that the mathematical structure of the word frequency distributions is of the compound Poisson type, and that Yule's hypothesis to that effect, which Simon thinks is not suitable, was correct.[3]

But Yule's accident hypothesis is a solution *a posteriori*. It was from the word frequency distribution as given that he constructed his hypothesis. It is, therefore, only a statistical solution, which will hardly satisfy the linguist. Yule's λ's, standing for the basic occurrence frequencies in groups of words, are not easy to interpret linguistically, and the frequency hypothesis of word occurrence has, therefore, its opponents among linguists.[10]

Even when the fact of statistical stability of such distributions is admitted, it is argued that they are not linguistically relevant, but are

only the outcome of general statistical properties of the material or physical features of language, not of language itself. To counteract this argument, we must keep in mind that although the samples from one text, or from different texts, combined into one overall distribution are statistically independent, they yet have a common underlying structure: the word code of the given language. This is the counterpart, in the plane of *expression*, to the concatenation of concepts, in the plane of *content*. And here would seem to be the reason why distributions with different origin can combine into one homogenous distribution of essentially the same type. Briefly, we may express this by saying that word frequency distributions are *statistically independent, but semiologically connected*.

We shall now show that the pattern of word connectivity in different samples of one text, or in samples from different texts, provides a truly linguistic interpretation of the peculiarity of the word distribution function, and the mechanism by which it arises. The full explanation would involve an exposition of the *Random Partitioning Function* (so-called by Yule) of vocabulary among samples or texts. However, so that this paper shall not be made inordinately long, the argument will only be developed in outline and the reader is referred for greater detail to G. Herdan, reference 11, chapters 3 and 17.

It is a characteristic property of the word code that if a text is divided into 2, 3 . . . n equal parts, or if equal-sized samples from 2, 3 . . . n different texts are analysed with regard to vocabulary, the number of words peculiar to one part or sample, or common to 2, 3 . . . n parts, can be calculated from the frequency distribution function resulting from the combination of all parts or samples, by a chance procedure called the 'random partitioning function'. This is the statistical way of arriving at those numbers. However, the dynamic process for the numbers of words to accumulate takes place, of course, during the composition of the text, and the accumulated word frequency distribution is only the statistical end result. We must always keep these two aspects, the statistical and the dynamic, well in mind. The fact that by random partitioning each part of the text has an equal share in the words belonging to the various basic frequency groups, accounts for the divisibility of the whole distribution into parts with similar frequency distributions.

The general pattern of the vocabulary connectivity between parts is as follows. The classes of one-sample and all-sample words are the most numerous classes, the former being the larger of the two; the average number of words peculiar to a combination of samples decreases as the number of samples included in the combination increases until a minimum in class frequency is reached for words of the middlemost class; from then on, the class frequencies rise again but in such a way that the classes have slightly smaller frequencies than the symmetrically corresponding classes on the other side of the distribution.

In particular, there are characteristic constants for certain classes or

groups of classes. The ratio of the sum of one-sample words plus all-sample words, i.e. of words peculiar to samples and common to all of them, to the vocabulary in each sample, is a constant for a given number of divisions. In other words, it depends only upon that number. For instance, for division of a word frequency distribution into three parts, the ratio referred to is about 70%, for division into four parts it is between 55% and 60%, for five parts about 50%, for nine parts somewhat over 30%, and so on.

Furthermore, whereas the one-sample words in the total distribution are about one-half of the total vocabulary (provided that the number of parts is not too small), indicating a one-to-one balance between one-sample and two-or-more-sample words together, there is also a balance between two-sample words and the rest, i.e. three-or-more-sample words, the ratio approaching $\frac{1}{2}$ as the number increases. In terms of the total vocabulary V, the two groups are about $\frac{1}{6}$ and $\frac{1}{3}$.

Illustrations. The relationships between one-sample words, two-sample words, and the total vocabulary are exemplified by Tables VII and VIII. The basic data for Table VII are in reference 11, chapter 17 and for Table VIII in reference 3.

TABLE VII

Partition of Greek New Testament, or of portions of it

	3 Parts	5 Parts	9 Parts (*Paul*)		9 Books (Whole *New Testament*)	
$\frac{V}{\text{One-sample words}}$	$\frac{901}{610}$ $= 1{\cdot}477$	$\frac{1271}{858}$ $= 1{\cdot}481$	$\frac{2648}{1415}$ $= 1{\cdot}871$		$\frac{5436}{2626}$ $= 2{\cdot}070$	
			Obs.	Theor.*	Obs.	Theor.*
$\frac{V}{\text{Two-sample words}}$	$\frac{901}{181}$ $= 4{\cdot}978$	$\frac{1271}{217}$ $= 5{\cdot}857$	$\frac{2648}{460{\cdot}8}$ $= 5{\cdot}747$	$\frac{2648}{475{\cdot}2}$ $= 5{\cdot}572$	$\frac{5436}{978{\cdot}5}$ $= 5{\cdot}555$	$\frac{5436}{918}$ $= 5{\cdot}922$

* Calculated by the Random Partitioning Function.

TABLE VIII

Partition into four parts

	Sampling Experiment on Macaulay's *Essay on Bacon*	Four of Macaulay's *Essays*	Four Works by Bunyan
$\frac{V}{\text{One-sample words}}$	$\frac{2048}{1138} = 1{\cdot}800$	$\frac{3543}{1698} = 2{\cdot}087$	$\frac{2246}{1250} = 1{\cdot}797$
$\frac{V}{\text{Two-sample words}}$	$\frac{2048}{405} = 5{\cdot}057$	$\frac{3543}{772} = 4{\cdot}589$	$\frac{2246}{446} = 5{\cdot}036$

The value of the ratios in question are reasonably close to the theoretical values for the word frequency distribution, namely 2 and 6 respectively, considering that they cannot be expected to equal these

values exactly, and show the corresponding relations for the word frequency distribution to be only the reflection of vocabulary connectivity between the parts of a text, as a consequence of the properties of the word code. Although the relation between the word frequency distribution and the vocabulary partition as given by the random partitioning function is such that the *hapax legomena* are the statistical source for most of the one-sample words, and the *dislegomena* the source for most of the two-sample words (though not quite to the extent as in the first case), yet it is only when the division is continued into smaller and smaller parts that a very close approach of the corresponding classes of words, in both the frequency distribution and the random partitioning function, can be expected; and even then only for very small values of i.

In the limit, when the parts are so small as to comprise only the words peculiar to a part and the words common to two parts, these words will correspond, by and large, to the *hapax legomena* and the *dislegomena* of the word frequency distribution, respectively, the former correspondence being closer than the latter because, as observed above, the *hapax legomena* have one-sample words as their only source, whereas the *dislegomena*, although having two-sample words as their main source, may also derive from some one-sample words. This is, of course, how the relation can be described in a dynamical way, that is during the accumulation of the word frequency distribution. Statistically, when we should start from the word frequency distribution, we should have to describe the *hapax legomena* and *dislegomena* as the respective sources for the one-sample and two-sample words.

This replaces Simon's idea of an abstract probability mechanism as accounting for the characteristic ratios for small values of i in the word frequency distribution, by *vocabulary connectivity* as the result of the concatenation of concepts or, as we say today, the structure of meaning, which is in full agreement with de Saussure's idea of language as 'un système où tout se tient'.

The essence of this way of accounting for the peculiar features of word frequency distributions is that in place of Simon's stochastic model, we have used the theorem that word frequency distributions, though statistically independent, are yet semiologically connected.

BIBLIOGRAPHY

[1] Simon, H. A. (1955). 'On a class of skew distribution functions', *Biometrika*, **42**, 425.

[2] Morgenthaler, R. (1958). *Statistik des Neutestamentlichen Wortschatzes*. Gotthelf-Verlag, Zürich-Frankfurt A.M.

[3] Yule, G. Udny (1944). *The Statistical Study of Literary Vocabulary*. Cambridge University Press.

[4] Josselson, H. H. (1953). *The Russian Word Count* (Para. 5 by B. Epstein). Wayne State Univ. Press, Detroit.

[5] Herdan, G. (1956). *Language as Choice and Chance*, Chapter 3, pp. 23-44. P. Noordhoff, Groningen.

[6] Thorndike, E. L. and Lorge, I. (1952). *The Teacher's Word Book of 30,000 Words*, 2nd ed. Columbia University Press.
[7] Eldridge, R. C. (1911). *Six Thousand Common English Words*. Privately Printed at Niagara Falls, New York.
[8] Van der Beke, G. E. (1929). *French Word Book*. The Macmillan Company, New York.
[9] Feller, W. (1957). *An Introduction to Probability Theory and its Applications, Vol. I*, 2nd ed., pp. 268ff. Wiley, New York.
[10] Halle, M. (1958). Review in *Kratylos* III, Part I, p. 20.
[11] Herdan, G. (1960). *Type-Token Mathematics. A Textbook of Mathematical Linguistics*. Mouton & Co., The Hague.

AN APPLICATION OF NON-PARAMETRIC STATISTICS TO TRUNCATED SELECTION

G. M. TALLIS
C.S.I.R.O., Division of Animal Genetics, McMaster Laboratory, Glebe, New South Wales

In this article Dr Tallis describes a procedure which saves time and labour when a proportion of a large population has to be selected on the basis of some measurement. Details of the calculations necessary are indicated for the general case, and for the particular cases likely to be of practical interest the results of these calculations are presented in a table to facilitate application.

Introduction

In animal and plant breeding, selection for a single character is frequently effected by truncation. More specifically, if x is a particular metrical character of interest, then the usual procedure is to discard all $x<b$, where b is a constant chosen in such a manner that the proportion of individuals retained is p, say.

Unfortunately, the point of truncation, b, is not usually known until all measurements are made, and hence all individuals must be tagged and identified with their particular scores. If the size of the group is large the latter operation may be expensive and inconvenient. Therefore in some cases it may be desirable to estimate b from a random sample taken from the group prior to measurement, so that individuals can be selected, at the time of measuring, on the basis of the estimate. It is the purpose of this paper to develop a method to accomplish this and to indicate the expected efficiency associated with the technique.

A good example of a situation where some prior estimate of b is desirable occurs in large flocks of sheep at shearing time. Recently, many commercial wool growers in Australia have been selecting their replacements on fleece weights and usually selection is completed at the conclusion of shearing. If the grower requires to save the best sheep available for replacements and he needs a proportion p of the total replacements flock, then obviously he is faced with the problem of truncated selection discussed above. In these circumstances the usual procedure is to tag each sheep and identify it with its fleece weight so that the best $100p\%$ animals may be determined from the records and sorted out of the main flock after shearing. If the flock is large, the amount of work involved in this operation is considerable and the methods of this paper have been devised with the aim of reducing the amount of labour of such selection programmes, while keeping in mind the limited facilities for performing computations in the field.

Methods

Let a sample of size n be drawn at random from a group which is to undergo truncated selection for the character x, and let the proportion to be retained be p. If the sample is ranked in descending order of x, then the n observations may be indicated by $x_1, x_2, \ldots x_n$, where subscripts refer to particular rankings. For instance, x_s is the measurement corresponding to the sth member of the ranked sample. Now, if $f(x)$, a continuous density function, is used to approximate the distribution of x in the group from which selection is to be made, then it is possible to estimate the percentage points of $f(x)$ from the ordered sample, $x_1, x_2, \ldots x_n$. In fact it is easily verified that the area of $f(x)$ lying to the left of x_s, w say, is distributed as a beta density with parameters s and $N-s$, where $N=n+1$. This result is independent of the type of density function, $f(x)$.

The beta density function may now be used to solve the following problems:

(*a*) What value of s, $s=l$ say, satisfies the equation $\Pr(w>p)=1-\alpha$, given a fixed n. This amounts to solving

$$\frac{1}{B(s, N-s)}\int_p^1 w^{s-1}(1-w)^{N-s-1}dw = 1-\alpha \qquad \ldots . (1)$$

for s.

(*b*) Once $s=l$ is determined, what value of p, p_1 say, satisfies the equation $\Pr(w<p_1)=1-\beta$ for fixed n and l. The required equation is

$$\frac{1}{B(l, N-l)}\int_0^{p_1} w^{l-1}(1-w)^{N-l-1}dw = 1-\beta \qquad \ldots . (2)$$

which must be solved for p_1.

The details of the solution of these equations are rather uninteresting. Briefly, (1) is transformed to Fisher's z distribution by the change of variable

$$z = \tfrac{1}{2}\ln\left\{\frac{(1-q)w}{q(1-w)}\right\} \qquad q = \frac{s}{N}$$

It is found that under this transformation z is distributed with parameters $n_1=2s$ and $n_2=2(N-s)$. From the asymptotic expansion of the z distribution presented by Fisher and Cornish (1960) it is possible to find an $s=l$ such that

$$\Pr\left[2z>\ln\left\{\frac{(1-q_1)p}{q_1(1-p)}\right\}\right] \simeq 1-\alpha \qquad q_1 = l/N$$

The latter equation must be solved by iterative techniques. Once l is found, it is a simple matter to calculate a z_1, such that $\Pr(z<z_1)=1-\beta$. The required quantity p_1 satisfying (2) is then obtained immediately from the inverse transformation

$$p_1 = \frac{q_1 e^{2z_1}}{1 + q_1(e^{2z_1} - 1)}$$

Equations (1) and (2) have been solved for various values of n and p for the case when $\alpha - \beta = 0{\cdot}05$ and the results of these calculations are recorded in Table I (p. 82). Next to the entries for p_1 there is a column for the 'efficiency', E_1, the derivation of which will now be given.

From equations (1) and (2), it is clear that Table I has been constructed so that, given the desired proportion of individuals to be retained, p, and the sample size, n, then if x_l is used as the estimate of the true truncation point b, then $\Pr(p < w < p_1) \simeq 0{\cdot}90$, where w is the actual proportion retained. Now, since $\mathrm{Av}(w) = l/N = q_1$, it is clear that instead of the desired proportion p being saved, on the average q_1 will be retained. Moreover, $q_1 > p$ and hence in order to finally obtain p selected individuals, on the average $(q_1 - p)$ will have to be discarded at random from the selected group. This procedure must obviously result in a certain loss of efficiency. In order to obtain some idea of the magnitude of this loss, it is convenient to assume $\mathrm{f}(x)$ approximately normal and to compare the selection differentials associated with p and q_1. (The selection differential is defined as the difference between the means of the selected and unselected groups.) If Δ and Δ_1 represent the selection differentials associated with p and q_1 respectively, then for the present purpose efficiency is defined as

$$E = \frac{\Delta_1}{\Delta} = \frac{\mathrm{Z}(q_1)p}{\mathrm{Z}(p)q_1}$$

where $\mathrm{Z}(p)$ and $\mathrm{Z}(q_1)$ are the ordinates of the normal curve associated with proportions p and q_1. Of course, if $\mathrm{f}(x)$ deviates from normality the values of E_1 in Table I may not necessarily apply. However, these figures are probably sufficient to obtain a good indication of the way the efficiency at the average truncation percentage changes with varying p and n. In order to avoid undue loss of efficiency it seems desirable to tag all sheep in the sample so that these sheep may also be selected on the basis of the estimate of b.

For example, suppose 60% of a group of individuals is to be retained, and an efficiency of 80% is considered tolerable. Then $p = 0{\cdot}6$, $E_1 = 80$, and from the Table, $n = 75$, $l = 52$. Thus 75 individuals would need to be measured and ranked, the truncation point being the measurement of the individual of 52nd rank. The 90% confidence interval for w is 0·6 to 0·768, that is, truncation at the 52nd rank would result in retaining from 60% to 77% of the individuals. The selected group may be adjusted later to the required size by discarding individuals at random. When the odd one in twenty chance occurs and $w < 0{\cdot}6$, then the additional numbers required are obtained at random from the unselected individuals; here again some loss of efficiency may occur.

Alternative Procedure

In order to complete the required truncation by the above method, it is only necessary to divide the population into two groups. The top group is retained and is adjusted to the required p value by discarding individuals at random. However, it can be seen that, when n is small, this procedure may result in a considerable loss of efficiency.

In order to overcome this, a slightly more complicated method of selection can be adopted. If we solve the equation,

$$\frac{1}{\mathrm{B}(s, N-s)} \int_0^p w^{s-1}(1-w)^{N-s-1}\mathrm{d}w = 1-\gamma \qquad \ldots . (3)$$

for s, we can find an $s=u$, say, such that if truncation is effected at x_u then $\Pr(w<p) = 1-\gamma$. If the results of (1) and (3) are combined it can be seen that if the true percentage point of $\mathrm{f}(x)$ corresponding to p is x_p, then $\Pr(x_l<x_p<x_u) = 1-\alpha-\gamma$. Equation (3) has been solved for various combinations of n and p with $\gamma = 0{\cdot}05$ and the values of u are also given in Table I.

Note: In order to solve (3) for s with $\gamma = 0{\cdot}05$, no calculations are necessary. By the change of variable $y = 1-w$ in (3) we find that

$$\frac{1}{\mathrm{B}(u, N-u)} \int_0^p w^{u-1}(1-w)^{N-u-1}\mathrm{d}w$$

$$= \frac{1}{\mathrm{B}(N-u, u)} \int_{1-p}^1 y^{N-u-1}(1-y)^{u-1}\mathrm{d}y \simeq 1-\gamma$$

The second integral is in the form of (1) with $N-u=l$ and $1-p$ replacing p. Therefore, for a particular n and p, u is found by $u = N-l$, where l is the value which has already been calculated for n, $1-p$.

The efficiency, E_2, is calculated on the same assumptions associated with E_1. If all individuals with x values greater than x_u are retained, then $\mathrm{Av}(w) = q_2 = u/N$. Now $p>q_2$, and hence, on the average, it will be necessary to obtain the required proportion p by adding $(p-q_2)$ individuals from the group where $x_l<x<x_u$. On the assumption of normality, the expected selection differential of the group selected in this manner is

$$\Delta_2 = \frac{1}{p}\left\{\mathrm{Z}(q_2) + \frac{(p-q_2)[\mathrm{Z}(q_1)-\mathrm{Z}(q_2)]}{q_1-q_2}\right\}$$

and

$$E_2 = \Delta_2/\Delta = \frac{(p-q_2)\mathrm{Z}(q_1) - (p-q_1)\mathrm{Z}(q_2)}{q_1-q_2}$$

As previously, the practical application of this procedure requires that p and E_2 be set in advance so that the appropriate n can be found from the table. A random sample is then drawn from which x_l and x_u are obtained. The remainder of the population may now be measured and put into three groups:

(A) individuals with measurements less than x_l (discards),
(B) individuals with measurements less than x_u and greater than x_l (reserves),
(C) individuals with measurements greater than x_u (retained).

If this is done, then l and u have been determined in such a way that nine times out of ten some individuals from group B will have to be taken at random and added to group C to bring the numbers of selected individuals to requirement. In the odd case when this is not so, adjustment must be made by either discarding animals at random from C when C is too large, or by adding animals at random from group A when the combined numbers of groups B and C are too small.

As an illustration, take the sample of 75 individuals previously considered. By retaining all above the individual 38th in rank, and putting those between the 38th and 52nd rank temporarily in reserve for later adjustment, the efficiency of the procedure is increased from 80% to 97%.

The choice between the two methods of selection presented here must be made on practical grounds. If it is inconvenient to break the population into three groups, then the first method must be used with its associated lower efficiency. However, whenever three groups can be handled satisfactorily, the high efficiency of the alternative procedure makes it the obvious choice.

In the particular example of selection for fleece weight mentioned in the introduction, a great reduction in clerical and physical labour can be expected by the use of either methods. The tedious procedure of writing 'roll call' lists and the even more unpleasant task of roll calling large flocks are avoided and this may represent a saving of a considerable number of man-days. The choice of which estimation procedure to use may be determined by the type of facilities and labour available.

As a final point, it is interesting to consider the relation of n to the efficiencies E_1 and E_2. These efficiency ratios simply reflect changes in the average selection differential as a result of taking larger samples; in the particular case of sheep selection, by the application of rather elementary genetical arguments, it is possible to approximate the annual loss of income due to the use of any selection method for any n from the efficiencies listed in Table I. Since it is suggested that individuals in the sample be tagged and identified with their scores, it would be possible to draw up cost curves by plotting economic gains against n. It would only be necessary to cost the labour of tagging and sorting 25, 50, 75, . . . sheep, and to take into account annual loss of income due to the statistical inefficiency of the method in order to draw up such a cost chart. From the graph it would then be possible to determine the optimum selection procedure to adopt.

Acknowledgements

I am indebted to Miss Elaine Smith for her assistance with the pre-

paration of Table I, and should like to acknowledge the helpful advice of Mr H. Weiler of the Division of Mathematical Statistics.

REFERENCE

FISHER, R. A. and CORNISH, E. A. (1960). 'The percentile points of distributions having known cumulants', *Technometrics*, **2**, 209.

TABLE I

Values of l, u, p_1, E_1, and E_2 for various combinations of p and n

p	0·1					0·2					0·3				
n	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2
25	6	1	0·375	75·1	87·5	9	2	0·504	76·2	88·2	12	4	0·621	74·2	90·1
50	9	2	0·270	83·6	92·3	15	6	0·403	83·7	95·1	21	10	0·526	81·5	95·2
75	12	4	0·231	87·0	95·7	21	10	0·364	86·5	96·7	30	17	0·488	84·1	97·1
100	16	6	0·221	87·0	95·9	27	14	0·342	88·0	97·4	38	23	0·456	87·1	97·7
150	22	10	0·196	89·4	97·5	39	23	0·318	89·5	98·1	55	36	0·430	89·0	98·3
200	28	14	0·181	90·8	98·1	50	31	0·300	91·0	98·6	71	50	0·409	90·8	98·9
500	62	40	0·149	94·1	99·4	115	86	0·261	94·3	99·6	167	134	0·368	94·2	99·6
1000	116	85	0·133	95·9	99·7	221	180	0·243	96·0	99·8	324	277	0·348	95·7	99·8
p	0·4					0·5					0·6				
n	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2
25	15	7	0·730	70·3	92·2	17	9	0·798	70·7	92·5	19	11	0·860	70·1	92·2
50	26	15	0·624	81·0	96·1	31	20	0·717	79·2	96·3	36	25	0·805	75·8	96·1
75	38	24	0·594	82·6	97·2	45	31	0·683	82·2	97·3	52	38	0·768	80·8	97·2
100	49	33	0·567	85·1	97·9	59	42	0·664	83·7	97·8	68	52	0·748	83·3	97·9
150	70	51	0·530	88·7	98·7	86	65	0·635	86·4	98·5	100	81	0·724	85·8	98·7
200	92	69	0·516	89·7	98·9	112	89	0·614	88·9	99·0	132	109	0·711	86·9	98·9
500	219	182	0·474	93·3	99·5	269	232	0·573	92·7	99·6	318	282	0·670	91·9	99·6
1000	426	375	0·451	95·3	99·8	526	475	0·551	94·9	99·8	626	575	0·650	94·2	99·8
p	0·7					0·8					0·9				
n	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2	l	u	p_1	E_1	E_2
25	22	14	0·943	56·5	90·1	24	17	0·985	44·7	88·2	25	20	0·998	44·1	87·5
50	41	30	0·887	69·3	95·2	45	36	0·946	64·1	95·1	49	42	0·993	45·1	92·3
75	59	46	0·850	77·6	97·5	66	55	0·926	70·4	96·7	72	64	0·982	58·5	95·7
100	78	63	0·837	78·8	97·7	87	74	0·913	73·5	97·4	95	85	0·974	64·1	95·9
150	115	96	0·816	81·8	98·3	128	112	0·893	79·3	98·1	141	129	0·963	70·5	97·5
200	151	130	0·800	85·0	98·9	170	151	0·886	80·1	98·6	187	173	0·957	74·0	98·1
500	367	334	0·764	90·6	99·6	415	386	0·855	88·0	99·6	461	439	0·939	82·9	99·4
1000	724	677	0·746	93·3	99·8	821	780	0·840	91·4	99·8	916	885	0·929	87·2	99·7

p = proportion of individuals to be retained.
n = number of individuals which are measured, their measurements then being ranked in order of merit as $x_1, x_2, \ldots x_u, \ldots x_l, \ldots x_n$.
l = rank whose measurement (x_l) is used to split the population into two groups.
u = rank whose measurement (x_u) is used to split the population into three groups. In this case u and l define the limits of the reserve group.
p_1 = upper 95% limit of proportion actually retained using the two group method.
E_1 = efficiency (%) of two group method at average truncation point.
E_2 = efficiency (%) of three group method at average truncation point.

CONFOUNDING IN 3^3 FACTORIAL EXPERIMENTS IN NINE BLOCKS

D. WILKIE
United Kingdom Atomic Energy Authority
Windscale and Calder Works

In this article Dr Wilkie compares two experiments, one in which interactions only are confounded and one in which two main effects are confounded. Although it is not customary to confound main effects there are instances where it is advantageous to do so.

Introduction

In a singly replicated 3^3 confounded factorial experiment in nine blocks of three treatments it is customary to leave the six main effects unconfounded. It can be shown that for quantitative factors having equally spaced levels all of the remaining twenty interactions are confounded. If, however, the main effects of one factor are confounded, there are as many as twelve unconfounded effects, consisting of the remaining four main effects and all eight two-factor interactions involving the confounded factor. Thus if the main effects of a factor are known to be, or expected to be, unimportant—something that it may not be possible to say of its two-factor interactions—this design is better than the normal one.

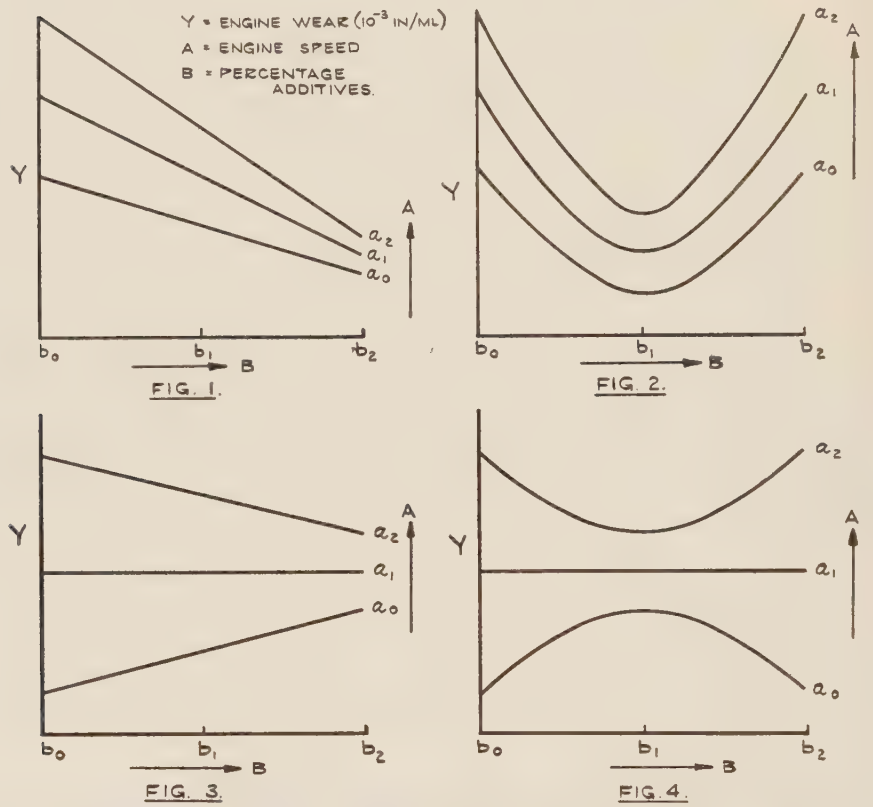
Main effects can be considered unimportant when they are not statistically significant or when, although statistically significant, they are of no interest to the experimenter because they will in no way affect the conclusions to be drawn from the experiment.

Into the first category come some chemical reactions in which, because of the presence of an inhibitor and a catalyst at the same time, inhibition takes place at, say, the low levels of another factor while catalysis takes place at the high levels. For example, adsorbed gas may act as an inhibitor at low temperatures but as temperature is increased the gas is evolved and the catalytic effect becomes predominant. There is therefore a strong interaction between temperature and the factor causing the introduction of the catalyst and inhibitor, but only a small or zero main effect of the latter factor.

Into the second category comes the type of experiment in which the main effects of the factor that are not of interest are known to be positive or negative for all levels of the other factors, and the factor is one over which the experimenter can exercise no control outside the experiment. The factor is included in the experiment because of interest in its possible interactions with other factors.

For example, a petrol manufacturer wishing to determine the best

percentage of additives for reducing engine wear may want to include engine speed as a factor, it being known already, let us assume, that an increase in engine speed always leads to greater engine wear. Not only the main effects of additives but also the interactions between additives and engine speed will influence his decision on the most suitable amount of additives to adopt. This is demonstrated in Figs. 1-4. For response curves as shown in Fig. 1 and Fig. 3, b_2 percentage additives would be the most satisfactory whereas for those of Fig. 2 and Fig. 4, b_1 percentage additives would be best, based on the criteria of least



FIGS. 1-4. Hypothetical example of a situation in which a factor is of interest only because of its interaction with another factor.

engine wear and least variation. The change in decision from 1 to 2 is brought about by the different main effects of B whereas the change in decision from 3 to 4 is caused by the different interactions between additives and engine speed. The decisions are not in any way affected by the magnitudes of the main effects of engine speed—as long as they are known to be positive.

The Designs

If three factors denoted by A, B, and C each at three levels 0, 1, and 2 are arranged in nine blocks of three, the 27 treatments can be ar-

ranged as in Table I(a) so that no main effects are confounded, or as in Table I(b) so that the main effects of A are confounded.

TABLE I

Allocation of treatments to blocks

Design	Block								
	1	2	3	4	5	6	7	8	9
(a) Main effects not confounded	000	210	120	100	010	220	200	110	020
	211	121	001	011	221	101	111	021	201
	122	002	212	222	102	012	022	202	112
(b) Main effects of one factor, A, confounded	000	010	020	100	110	120	200	210	220
	011	021	001	111	121	101	211	221	201
	022	002	012	122	102	112	222	202	212

The numerals in each triple figure refer to the levels of the three factors A, B, and C in that order.

Contribution of Block Levels to Total Effects

The mean responses for the nine blocks may be denoted by the letters r, s, t, u, v, w, x, y, z respectively. These may be analysed as if they were the actual responses for all three treatments in each block. In this way the contributions of the blocks to the total effects can be determined and it can be seen directly what effects are confounded. In the analysis of experimental results the individual results and not the block means would, of course, be analysed (as in the example of Table VI).

The analysis is similar to the standard method of Yates[1] for 2^n experiments but with the extension to three levels.[2, 3] By this means, the linear and quadratic components of the main effects and interactions are calculated.

The analysis is carried out as in Table II(a) for the design with main effects unconfounded and Table II(b) for the design with main effects of one factor confounded. In both cases the responses are arranged in standard order of treatment, that is, from the lowest to the highest level of the first factor, first while the other factors remain at the lowest level, secondly when the second factor alone moves to the middle level, and thirdly when the second factor is at its highest level. This order is repeated for the middle and finally the highest levels of the third factor. The column of responses is divided into groups of three and operated on by the table of operators (Table III) which are the orthogonal polynomials for $n = 3$ (see, for example, reference 4). The first column of operators indicates that the responses in each group of three should be added together to provide the top one-third of column 1 of Table II (a) and (b), the second column of operators indicates that in each group of three responses the first should be subtracted from the third to provide the values for the middle one-third of column 1, and the third column of operators indicates that in each group of three responses twice the second should be subtracted from the sum of the other two to provide the values for the bottom one-third of column 1.

Column 1 is treated in the same way as the column of responses to yield column 2, which in turn yields column 3. Column 3 shows the contributions made by the blocks to the total effects, which are indicated by the symbols in the final column (e.g. the symbol AB^2 indicates the interaction of the linear effect of A with the quadratic effect of B).

TABLE II

Contribution of blocks to total effects

(a) *Main effects unconfounded*

Treatment *a*	*b*	*c*	Mean Response	Column 1	Column 2	Contribution of Blocks, Column 3	Symbol
0	0	0	r	$r+u+x$	$r+s+t+u+v+w+x+y+z$	$3(r+s+t+u+v+w+x+y+z)$	Mean
1	0	0	u	$v+y+s$	$r+s+t+u+v+w+x+y+z$	0	A
2	0	0	x	$z+t+w$	$r+s+t+u+v+w+x+y+z$	0	A^2
0	1	0	v	$t+w+z$	$x-r+s-v+w-z$	0	B
1	1	0	y	$u+x+r$	$z-t+r-u+v-y$	$r+s+t+u+v+w-2x-2y-2z$	AB
2	1	0	s	$y+s+v$	$y-s+t-w+u-x$	$3(u+v+w-r-s-t)$	A^2B
0	2	0	z	$s+v+y$	$r+x-2u+v+s-2y+z+w-2t$	0	B^2
1	2	0	t	$w+z+t$	$t+z-2w+u+r-2x+y+v-2s$	$3(u+v+w-r-s-t)$	AB^2
2	2	0	w	$x+r+u$	$s+y-2v+w+t-2z+x+u-2r$	$3(2x+2y+2z-r-s-t-u-v-w)$	A^2B^2
0	0	1	t	$x-r$	$z+t+w-r-u-x$	0	C
1	0	1	w	$s-v$	$y+s+v-t-w-z$	$v+r+t+u+y+z-2s-2w-2x$	AC
2	0	1	z	$w-z$	$x+r+u-s-v-y$	$3(t+u+y-r-v-z)$	A^2C
0	1	1	u	$z-t$	$w-z-x+r$	$2r+2u+2x-s-t-v-w-y-z$	BC
1	1	1	x	$r-u$	$v-y-z+t$	$s+u+z-r-y-w$	ABC
2	1	1	r	$v-y$	$u-x-y+s$	$2t+2v+2x-r-s-u-w-y-z$	A^2BC
0	2	1	y	$y-s$	$z+w-2t-r-x+2u$	$3(s+v+y-w-z-t)$	B^2C
1	2	1	s	$t-w$	$y+v-2s-t-z+2w$	$r+s+u+w+y+z-2t-2v-2x$	AB^2C
2	2	1	v	$u-x$	$x+u-2r-s-y+2v$	$3(s+u+z-r-w-y)$	A^2B^2C
0	0	2	s	$r+x-2u$	$z+t+w+r+u+x-2v-2y-2s$	0	C^2
1	0	2	v	$v+s-2y$	$y+s+v+t+w+z-2u-2x-2r$	$3(t+u+y-r-v-z)$	AC^2
2	0	2	y	$z+w-2t$	$x+r+u+s+v+y-2w-2z-2t$	$3(2s+2w+2x-r-t-u-v-y-z)$	A^2C^2
0	1	2	w	$t+z-2w$	$w-z+x-r-2s+2v$	$3(t+w+z-s-v-y)$	BC^2
1	1	2	z	$u+r-2x$	$v-y+z-t-2r+2u$	$r+s+u+w+y+z-2v-2t-2x$	ABC^2
2	1	2	t	$y+v-2s$	$u-x+y-s-2t+2w$	$3(s+u+z-r-w-y)$	A^2BC^2
0	2	2	x	$s+y-2v$	$z+w-2t+r+x-2u-2v-2s+4y$	$3(2r+2u+2x-s-t-v-w-y-z)$	B^2C^2
1	2	2	r	$w+t-2z$	$y+v-2s+t+z-2w-2u-2r+4x$	$3(r+w+y-s-u-z)$	AB^2C^2
2	2	2	u	$x+u-2r$	$x+u-2r+s+y-2v-2w-2t+4z$	$3(r+s+u+w+y+z-2t-2v-2x)$	$A^2B^2C^2$

In Table II(a) only the six main effects are unconfounded whereas in Table II(b) twelve effects are unconfounded, four of them being main effects, the other eight being all the two-factor interactions of A with B and C. The only linear two-factor interaction remaining confounded is *BC*.

TABLE II (*Contd.*)

Contribution of blocks to total effects

(b) *Main effects confounded*

Treatment *a*	*b*	*c*	Mean Response	Column 1	Column 2	Contribution of Blocks, Column 3	Symbol
0	0	0	r	$r+u+x$	$r+s+t+u+v+w+x+y+z$	$3(r+s+t+u+v+w+x+y+z)$	Mean
1	0	0	u	$s+v+y$	$r+s+t+u+v+w+x+y+z$	$3(x+y+z-r-s-t)$	A
2	0	0	x	$t+w+z$	$r+s+t+u+v+w+x+y+z$	$3(x+y+z+r+s+t-2u-2v-2w)$	A^2
0	1	0	s	$t+w+z$	$x+y+z-r-s-t$	0	B
1	1	0	v	$r+u+x$	$x+y+z-r-s-t$	0	AB
2	1	0	y	$s+v+y$	$x+y+z-r-s-t$	0	A^2B
0	2	0	t	$s+v+y$	$x+y+z+r+s+t-2(u+v+w)$	0	B^2
1	2	0	w	$t+w+z$	$x+y+z+r+s+t-2(u+v+w)$	0	AB^2
2	2	0	z	$r+u+x$	$x+y+z+r+s+t-2(u+v+w)$	0	A^2B^2
0	0	1	t	$x-r$	$t+w+z-r-u-x$	0	C
1	0	1	w	$y-s$	$s+v+y-t-w-z$	0	AC
2	0	1	z	$z-t$	$r+u+x-s-v-y$	0	A^2C
0	1	1	r	$z-t$	$r+z-t-x$	$2(r+u+x)-(s+v+y+t+w+z)$	BC
1	1	1	u	$x-r$	$t+y-s-z$	$-2r+s+t+2x-y-z$	ABC
2	1	1	x	$y-s$	$s+x-r-y$	$2(r+v+w+x)-s-t-y-z-4u$	A^2BC
0	2	1	s	$y-s$	$t+z-r-x+2(u-w)$	$3(s+v+y-t-w-z)$	B^2C
1	2	1	v	$z-t$	$s+y-t-z+2(w-v)$	$3(t+y-s-z)$	AB^2C
2	2	1	y	$x-r$	$r+x-s-y+2(v-u)$	$3(s+y+2w-t-z-2v)$	A^2B^2C
0	0	2	s	$x+r-2u$	$r+t+u+x+w+z-2(s+v+y)$	0	C^2
1	0	2	v	$y+s-2v$	$s+t+v+w+y+z-2(r+u+x)$	0	AC^2
2	0	2	y	$z+t-2w$	$r+s+u+v+x+y-2(t+w+z)$	0	A^2C^2
0	1	2	t	$z+t-2w$	$x+z-r-t-2(y-s)$	$3(t+w+z-s-v-y)$	BC^2
1	1	2	w	$x+r-2u$	$y+z-s-t-2(x-r)$	$3(s+z-t-y)$	ABC^2
2	1	2	z	$y+s-2v$	$x+y-r-s-2(z-t)$	$3(t+z-y-s+2v-2w)$	A^2BC^2
0	2	2	r	$y+s-2v$	$r+t+x+z-2(s+u+w+y)+4v$	$3(2r+2u+2x-s-v-y-t-w-z)$	B^2C^2
1	2	2	u	$z+t-2w$	$t+s+y+z-2(r+v+w+x)+4u$	$3(s+t+2x-y-z-2r)$	AB^2C^2
2	2	2	x	$x+r-2u$	$r+s+x+y-2(t+u+v+z)+4w$	$3(2r+2x+2v+2w-s-y-t-z-4u)$	$A^2B^2C^2$

Deconfounding

Some of the effects that are confounded can be combined to give groups that are free from block effects. The groups are therefore deconfounded.[3] Table IV shows one form that these groups can take. If the assumption is made that A^2B^2, B^2C^2, and A^2C^2 are negligible, the linear interactions AB, BC, and AC can be estimated in case (a). Two of these effects are obtained in case (b) without any assumption, BC

TABLE III

Operators for factors at three levels

	Levels		
	0	1	2
	Operators		
	1	−1	1
	1	0	−2
	1	1	1
Sums of squares	3	2	6

being the only one requiring it. It is reasonable to assume, and usually confirmed in practice, that the three-factor interactions are negligible, in which case six estimates of error are obtained in case (a) and four in case (b), all free from block contributions. The latter estimates can be included with the two unconfounded interactions A^2B^2 and A^2C^2 to give six estimates likewise.

TABLE IV

Deconfounding

Combinations of Total Effects that are Independent of Blocks	Interpretation
(a) Main effects not confounded	
$3AB + A^2B^2$	$3AB$
$3BC - B^2C^2$	$3BC$
$3AC + A^2C^2$	$3AC$
$A^2B - AB^2$	
$B^2C + BC^2$	
$A^2C - AC^2$	
$A^2BC + AB^2C$	Error
$A^2B^2C^2 - 3ABC^2$	Error
$AB^2C - ABC^2$	Error
$AB^2C^2 + A^2BC^2$	Error
$AB^2C^2 + A^2B^2C$	Error
$A^2B^2C - 3ABC$	Error
(b) Main effects confounded	
$3BC - B^2C^2$	BC
$B^2C + BC^2$	
$3ABC - AB^2C^2$	Error
$AB^2C + ABC^2$	Error
$A^2B^2C + A^2BC^2$	Error
$3A^2BC - A^2B^2C^2$	Error

The two-factor interactions involving the linear effect of one factor and the quadratic of another are confounded with each other in case

(a). If one of a pair is important it will be impossible to tell which. In case (b) only one such pair exists. Therefore the choice between the two designs depends on the importance of the linear-quadratic two-factor interactions on the one hand and the main effects of one factor on the other. As a rule the main effects are likely to be more important. When this is not so, the second design is to be preferred. In addition the latter design gives two of the linear two-factor interactions directly. The partitioning of the degrees of freedom is shown in Table V.

In both designs the unconfounded and deconfounded effects are unbiased by block differences, which would not be the case in an unconfounded factorial experiment. They are therefore estimated more precisely. At the same time the error will be smaller for it too has had the effect of blocks removed, so making the experiment more sensitive.

TABLE V

Partitioning of degrees of freedom in 3^3 factorial design with confounding

Source of Variation	Degrees of Freedom	
	Case (a)	Case (b)
Main effects	6	4
2-factor interactions:		
linear	3	3
linear quadratic	3	5
Blocks	8	8
Error	6	6
Totals	26	26

Example

A sample analysis of a confounded experiment in nine blocks of three with the main effects of A confounded is given in Table VI. These effects were expected to be non-significant. The analysis proceeds as if there were no confounding [Table VI (a)]. The method is the same as was outlined for Table II. The divisors used in calculating the mean squares and the mean effects are given by the products of the operators sums of squares corresponding to the power of the factor (Table III). Thus for the effect AB^2, that is, $A^1B^2C^0$, the divisor is $2\times6\times3=36$ and for A^2BC^2 it is $6\times2\times6=72$. The reason for this is indicated by the way in which the effects are derived. For example effect $A^1B^2C^0$ is given by $(-a_0+a_2)(b_0-2b_1+b_2)(c_0+c_1+c_2)$ that is by the treatment sum

$$\begin{aligned}&-a_0b_0c_0+a_2b_0c_0+2a_0b_1c_0-2a_2b_1c_0-a_0b_2c_0+a_2b_2c_0\\&-a_0b_0c_1+a_2b_0c_1+2a_0b_1c_1-2a_2b_1c_1-a_0b_2c_1+a_2b_2c_1\\&-a_0b_0c_2+a_2b_0c_2+2a_0b_1c_2-2a_2b_1c_2-a_0b_2c_2+a_2b_2c_2\end{aligned}$$

Each response has the population variance, estimated by S_2^2, say, so that the variance of $A^1B^2C^0$ is

$$(1^2+1^2+2^2+2^2+1^2+1^2+1^2+1^2+2^2+2^2+1^2+1^2+1^2+1^2+2^2+2^2+1^2+1^2)S^2_2 = 36S_2^2$$

Therefore, the divisor used in obtaining the mean square for AB^2 will be 36 since the expected value for the mean square, when it has no real component, is the population variance.

Checks on the arithmetic can be made at each stage of the calculation. The sums of the rows of the operators in Table III are 1, −1, and 3, so that the sum of the first numbers in every group of three responses minus the sum of the second numbers plus three times the sum of the third numbers will be equal to the total for the next column. This check applies to the next two columns as well. An overall check is given by the sum of squares of the responses which should equal, to within rounding-off errors, the sum of the mean squares.

TABLE VI

Analysis of 3^3 factorial experiment with main effects confounded

(a) *Ignoring confounding*

Treatment *a*	*b*	*c*	Response			Total Effect	Divisor *d*	Mean Square	Mean Effect	Symbol
0	0	0	8	45	552	3209	27	381 395***	118·9	Mean
1	0	0	14	183	1373	23	18	(29)		A
2	0	0	23	324	1284	−55	54	(56)		A^2
0	1	0	54	410	45	247	18	3 389***	13·7	B
1	1	0	60	468	27	−30	12	75*		AB
2	1	0	69	495	−49	−14	36	5·4		A^2B
0	2	0	101	482	9	−55	54	56*		B^2
1	2	0	107	437	−7	2	36	0·1		AB^2
2	2	0	116	365	−57	2	108	0·04		A^2B^2
0	0	1	129	15	279	732	18	29 768***	40·7	C
1	0	1	136	15	85	−94	12	736***	−7·8	AC
2	0	1	145	15	−117	−66	36	121**	−1·8	A^2C
0	1	1	151	16	0	−396	12	(13 068)		BC
1	1	1	157	9	−14	−16	8	(32)		ABC
2	1	1	160	2	−16	−6	24	(1.5)		A^2BC
0	2	1	163	−8	0	−30	36	(25)		B^2C
1	2	1	167	−17	−8	2	24	(0·2)		AB^2C
2	2	1	165	−24	−6	0	72	(0)		A^2B^2C
0	0	2	162	3	3	−910	54	15 335***	−16·9	C^2
1	0	2	166	3	−31	−58	36	93**	−1·6	AC^2
2	0	2	154	3	−27	−34	108	10·7		A^2C^2
0	1	2	151	2	0	−8	36	(1·8)		BC^2
1	1	2	152	−3	0	12	24	(6·0)		ABC^2
2	1	2	134	−6	2	10	72	(1·4)		A^2BC^2
0	2	2	130	−16	0	38	108	(13·4)		B^2C^2
1	2	2	129	−19	2	2	72	(0·1)		AB^2C^2
2	2	2	106	−22	0	−4	216	(0·1)		$A^2B^2C^2$
Sums			—	3177	3329	2503		—		
Check totals			3177	3329	2503	—		—		
Sums of squares			444 221	—	—	—		444 219		

The confounded mean squares are in parentheses.
Significance: 5% *; 1% **; 0·1% ***.

TABLE VI (*Contd.*)

(b) *Deconfounding certain interactions*

Symbol	Total Effect	Divisor for Mean Effect	Divisor for Mean Square	Mean Square	Mean Effect	Interpretation
$3BC-B^2C^2$	$-1188-38$ $=-1226$	3×12 $=36$	$9\times12+108$ $=216$	6959***	−34·1	*BC*
B^2C+BC^2	$-30-8=-38$		$36+36=72$	20·0		
$3ABC-AB^2C^2$	$-48-2=-50$		$9\times8+72$ $=144$	17·4		Error
AB^2C+ABC^2	$2+12=14$		$24+24=48$	4·1		Error
$A^2B^2C+A^2BC^2$	$0+10=10$		$72+72=144$	0·7		Error
$3A^2BC-A^2B^2C^2$	$-18+4=-14$		$9\times24+216$ $=432$	0·5		Error

(c) *Significance test*

Probability	$F_{1,6}$	FS_2^2	$F_{8,6}$	FS_2^2
5% *	5·99	33·5	4·15	22·2
1% **	13·75	77·0	8·10	45·4
0·1% ***	35·51	198·9	19·03	106·6

Deconfounding proceeds as in Table VI(b) in which the confounded total effects are combined as in Table IV(b). The divisors for the mean squares of the deconfounded groups are obtained from the divisors for the components. Thus the variance of

$$\begin{aligned}(3BC-B^2C^2) &= 9\times\text{variance of } BC+1\times\text{variance of } B^2C^2\\ &= 9\times\text{divisor for } BC\times S_2^2+1\times\text{divisor for } B^2C^2\times S_2^2\\ &= (9\times12+108)S_2^2 = 216S_2^2\end{aligned}$$

giving a divisor of 216 for this group. As *BC* is the only deconfounded effect that clearly needs testing and the mean square of the deconfounded group in which it appears is significant (see later for actual test) its mean effect requires to be calculated. Since B^2C^2 is assumed to be zero the divisor for the mean effect of *BC* is $3\times$divisor for *BC*. The effect of group B^2C+BC^2 is clearly small and later proves to be insignificant.

The mean square due to blocks is obtained from the mean squares of the confounded effects after removing any real components. Thus the real component of the total effect of $BC=-1226/3=-409$ while the confounded total effect is -396. The block contribution to the confounded total effect is therefore $-396-(-409)=13$ and the mean square for the block contribution is $13^2/12=14{\cdot}1$. This is added to the mean squares of all the other confounded effects, which are assumed to have no real components, to give a sum of squares of

$$29+56+14+32+2+25+0+0+2+6+1+13+0+0 = 180$$

based on the eight degrees of freedom ascribed to blocks. This leads to a block mean square of 22·5.

The variance of the observations is the average of the six mean squares that have been ascribed to error and has a value

$$S_2^2 = (0{\cdot}04 + 10{\cdot}7 + 17{\cdot}4 + 4{\cdot}1 + 0{\cdot}7 + 0{\cdot}5)/6$$
$$= 5{\cdot}6 \text{ with } \nu_2 = 6 \text{ degrees of freedom.}$$

It is used for testing the significance of the remaining mean squares using the variance ratio test $F = S_1^2/S_2^2$ where S_1^2 is the mean square for the effects based on 1 degree of freedom except in the case of the block mean square, which is based on 8 degrees of freedom. Thus any mean square greater than FS_2^2 [Table VI(c)], where F is the tabulated ratio (see, for example, reference 4) for the appropriate degrees of freedom and probability level, will be significant at this level. The main effects of C, the linear effect of B, the interaction AC and the deconfounded interaction BC are all highly significant, and the interactions A^2C and AC^2 are next in importance. The block mean square is just significant at the 5% level.

Conclusions

Confounding a pair of main effects in a 3^3 factorial experiment gives twelve unconfounded effects compared with six when interactions only are confounded.

The choice between the two designs depends on the relative importance of the two main effects on the one hand and the six linear-quadratic two-factor interactions on the other hand.

REFERENCES

[1] Yates, F. (1937). The Design and Analysis of Factorial Experiments. Commonwealth Bureau of Soils, Harpenden, England.

[2] Davies, O. L. (1956). *Design and Analysis of Industrial Experiments*, p. 366. Oliver and Boyd, Edinburgh and London.

[3] Eisenklam, P. (1953). 'Statistical methods in engineering experimentation', *Research*, **6**, 195 and 231.

[4] Fisher, R. A. and Yates, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research*. Oliver and Boyd, Edinburgh and London.

INTERVIEWER DIFFERENCES AMONG AUTOMOBILE PURCHASERS

ALLEN F. JUNG
Graduate School of Business, University of Chicago

In this article Professor Jung describes an investigation into the possible differences among the interviewers used in a pricing study. The results also throw some light on the effects of some relaxation of retail price maintenance.

Through the years some researchers have found considerable differences among the interviewers they employed for their studies. This has been especially true in pricing studies where prices were not posted or where discounts were commonly given from the posted price. (These pricing studies were conducted by research firms for clients and were never published.) In these cases, the interviewer had to obtain the price from a salesman. These differences among interviewers may have occurred because real differences existed or because the best possible practices in respect to selection and training of personnel had not been followed.

The present study was designed in part to gain information about differences in interviewers. The interviewers were to be carefully selected and to receive sufficient training for the pricing task to be performed. The main purpose was to determine whether dealers varied prices when different prospective buyers, using the same approach, contacted them.

The study found that differences among interviewers were relatively small. All interviewers did not obtain identical prices as there were considerable variations among the dealers contacted. Moreover, in all cases but two a given dealer did not quote the same price to each interviewer that contacted him, despite the fact that all interviewers used the same approach and were pricing an identical product. All prices were obtained by all prospective buyers on the same day.

Method

All interviewers obtained prices from the selected automobile dealers by contacting them personally. Each interviewer posed as a prospective purchaser, and used the same standardised approach and shopping procedure on every buying contact. Each acted as a knowledgeable buyer and gave evidence of familiarity with the automobile priced and with car buying procedures. The would-be purchaser was definitely in the market for this new car at the time of contact, and would make the final decision about its purchase within a couple of days. However, he was willing to wait for delivery of the exact car desired, rather than

buy one from stock with different equipment on it. Every contact was stated to be the first stop on the tour, because it was felt that previous price quotations might bias the price a salesman would offer. The car was to be purchased for cash and without a car taken in trade. At every dealer the interviewer obtained the salesman's card.

After the salesman quoted a price, the interviewer stated that this seemed high and appeared to be over what he had expected to pay. He then asked the salesman for the lowest price at which he could sell the car. The price received at this point, whether the same or lower than the original quotation, was recorded. The latter price is the one used in all computations.

Each interviewer received written instructions and several hours of personalised instruction about the approach to the dealer, the automobile to be priced, and ways of dealing with questions that salesmen might ask. Some automobile salesmen go to great lengths to qualify a buyer before quoting a realistic price. It is essential that the prospective purchaser should indicate that he is definitely in the market for a new car and give some indication that he is more interested in getting a 'good deal' than in visiting every dealer in town to find the single lowest price. If the salesman feels that the buyer is interested only in price, he may quote a 'low-ball', a price at which the dealer would refuse to deliver the car. The 'low-ball' price is designed to force the price-conscious buyer to return before consummating the purchase. Upon return, the salesman usually informs the buyer that a mistake was made in the price, or that the buyer misunderstood the model and/or accessories. Some dealers will write an order at the 'low-ball' price, but will inform the consumer and ask for additional money at time of delivery to complete the transaction.

Sample

The sample was selected from Falcon (Ford) factory authorised dealers located within the city limits of Chicago, Illinois. The same car was priced at every dealer. It was the Falcon 4-door sedan, with radio, heater, automatic transmission, windshield washers, and deluxe trim. This model was chosen because more 4-door Falcons have been sold than any other model, and a majority of these come equipped with deluxe trim. The radio, heater, and automatic transmission are very common accessories, and the windshield washers were included to lend credibility to the interviewer's intention to purchase a new car. Falcon offered only one each of the accessories above and this greatly simplified the task. Any solid colour was suitable. This car carried a factory suggested list price (the price that would be on the label attached at the factory) approximating $2373.

Twenty-eight of the twenty-nine Falcon dealers in Chicago were included in the study. Eight interviewers were employed, including myself, and every dealer was visited by two interviewers. Each interviewer was paired with every other interviewer once, which made a

total of twenty-eight teams. The dealers were assigned to the teams at random. Every dealer included was priced by one team.

One dealer had to be eliminated to use a balanced incomplete block design. The design used comprised 8 treatments in 28 incomplete blocks of size 2, and the blocks were not grouped into separate replications. This type of design is useful for situations which only require control of the variability among the incomplete blocks.[1]

The eight interviewers were composed of four men and four women. All contacts were completed satisfactorily on the same day, 4th May 1960, to keep any price differentials with respect to time at a minimum.

TABLE I

*Falcon prices quoted to the interviewers by the dealers**

Dealers	Interviewers								Dealer Means
	A	B	C	D	E	F	G	H	
1	100	125	—	—	—	—	—	—	113
2	235	—	95	—	—	—	—	—	165
3	50	—	—	30	—	—	—	—	40
4	133	—	—	—	—	80	—	—	107
5	50	—	—	—	30	—	—	—	40
6	25	—	—	—	—	—	88	—	57
7	140	—	—	—	—	—	—	150	145
8	—	41	50	—	—	—	—	—	46
9	—	180	—	195	—	—	—	—	188
10	—	65	—	—	75	—	—	—	70
11	—	50	—	—	—	100	—	—	75
12	—	100	—	—	—	—	96	—	98
13	—	170	—	—	—	—	—	150	160
14	—	—	75	95	—	—	—	—	85
15	—	—	25	—	—	55	—	—	40
16	—	—	132	—	50	—	—	—	91
17	—	—	145	—	—	—	—	96	121
18	—	—	100	—	—	—	152	—	126
19	—	—	—	99	235	—	—	—	167
20	—	—	—	100	—	100	—	—	100
21	—	—	—	50	—	—	50	—	50
22	—	—	—	35	—	—	—	50	43
23	—	—	—	—	150	163	—	—	157
24	—	—	—	—	135	—	150	—	143
25	—	—	—	—	70	—	—	138	104
26	—	—	—	—	—	50	100	—	75
27	—	—	—	—	—	75	—	65	70
28	—	—	—	—	—	—	100	89	95
Range	25–235	41–180	25–145	30–195	30–235	50–163	50–152	50–150	Grand Range 25–235
Interviewer Mean	105	104	89	86	106	89	105	105	Grand Interviewer Mean 99
Adjusted Interviewer Mean	111	99	87	84	99	94	114	102	

* $2000 has been subtracted from all prices.

Findings

The prices obtained at every dealer by each team and by each interviewer can be seen in Table I. The prices quoted by dealers

varied from $2025 to $2235. The mean prices at dealers ranged from a low of $2040 to a high of $2167. Mean prices per interviewer varied from $2086 to $2105. After adjustment for dealer variations, average interviewer prices ranged from $2084 to $2114. No adjusted interviewer average varied more than $15 from the mean, $2099.

In the analysis of variance all dealer and interviewer means were computed and the interviewer means were then adjusted for dealers contacted.[2] The various sums of squares are presented in Table II. There is no significant variation among interviewers, the mean square 768·1 being less than 1432·2, and one may conclude that interviewer differences are small. Since this is so, the unadjusted dealer means may be examined; the variation among dealers may be compared with the intra-dealer variance, and indicates significant variation among dealers. The efficiency of this design, using a balanced incomplete block, relative to the corresponding completely randomised design was increased approximately 40 per cent.

TABLE II

Analysis of variance

Source of Variation	DF	SS	MS
Total Uncorrected	56	688 134	—
Correction for Mean	1	546 483	—
Total	55	141 651	—
Interviewers (Eliminating Dealers)	7	5 377	768·1
Dealers (Ignoring Interviewers)	27	106 197·4	—
Intra Dealers	21	30 076·6	1432·2

As might be expected, since there were small differences among interviewers, the male interviewers did not obtain prices substantially below their female counterparts. The adjusted mean for men was $2098 compared with $2100 for the women. The natural business acumen of the men and the beauty and charm of the ladies must be considered offsetting factors as far as obtaining lower automobile prices.

Three interviewers, using three different approaches, contacted all Ford and Chevrolet dealers in Chicago in February, 1959.[3] Each interviewer mastered an approach and used it on every dealer contacted. They were quoted a wide variety of prices, and there were considerable price differences among dealers. Dealers did not systematically vary prices according to the approach used, but at a given dealer wide variations were found.

Conclusions

There were relatively slight differences among interviewers in an area where price variations have been considerable. Some researchers, in studies of this type, have found considerable differences among interviewers. In these studies, the interviewers were often relatively

untrained. This suggests that the relatively small differences found here might be attributed to more thorough training of the interviewers. However, interviewer differences can be adjusted for and the extra expense involved in additional training might not be justified in commercial research.

These findings about interviewer differences can be applied only to a very limited area of pricing. Where posted prices are market prices, there should be only small, if any, differences found among the individuals obtaining these prices.

One might wonder how the fact that interviewer differences are small can be reconciled with the widely different prices obtained at a specific dealer. Evidently different prices are offered by various salesmen at a particular dealer. Each salesman probably makes his own estimate of what price is necessary to close a sale with a particular purchaser. There were four instances in which both members of a team contacted the same salesman. The differences in the prices quoted by the same salesman were not substantial ($0, $15, $15, and $20).

When buying a car the consumer should realise that there are additional variations besides those among dealers. The variations in prices offered by different salesmen at a particular dealer suggest that here we have an important cause of the saving that results from contacting more dealers; this saving would not be so great if each dealer contacted always adhered to his own fixed price for a given car. Reference 3 contains a table that estimates the expected savings in dollars from contacting additional dealers.

REFERENCES

[1] Federer, W. T. (1955). *Experimental Design*, p. 416. Macmillan, New York.
[2] *Ibid.*, pp. 416-422.
[3] Jung, A. F. (1959). 'Price variations among automobile dealers in Chicago, Illinois', *Journal of Business*, **32**(4), 315.

THE ANALYSIS OF A $3^2 \times 2^2$ FACTORIAL EXPERIMENT WITH CONFOUNDING

R. H. E. INKSON
Macaulay Institute for Soil Research, Aberdeen

This article explains and discusses alternative methods of analysis of the results of an agricultural experiment.

Introduction

A series of factorial experiments, commenced in 1948 by the Macaulay Institute for Soil Research, Aberdeen, was designed to provide information on the responses to and the interactions between nitrogen, phosphorus, potassium, and farmyard manure applied to a number of crops grown in Scotland. The design was based on the methods described by Yates[4] for a $3^2 \times 2$ experiment in blocks of six plots. Since more importance was attached to determining the effect of dung on the responses to nitrogen, phosphate, and potash than to obtaining a precise measure of its effect on yield, this factor was completely confounded with blocks. This procedure simplified the practical details of applying the dung. The treatments consisted of all the thirty-six combinations of the four factors:

Nitrogen	0, 30, 60 lb N per acre (n_0, n_1, n_2)
Phosphate	0, 40, 80 lb P_2O_5 per acre (p_0, p_1, p_2)
Potash	0, 60 lb K_2O per acre (k_0, k_1)
Dung	0, 15 tons dung per acre (d_0, d_1)

The design, layout and yields for one experiment of the series are shown in Table I where 2200 denotes the treatment combination $n_2p_2k_0d_0$, 2010 denotes $n_2p_0k_1d_0$, etc. Following the notation used by Yates[4] in defining the diagonal sets of combinations of n and p, consider the I totals:

$$I_1 = n_0p_0 + n_1p_1 + n_2p_2 \qquad I_2 = n_0p_2 + n_1p_0 + n_2p_1$$
$$I_3 = n_0p_1 + n_1p_2 + n_2p_0$$

The sum of squares for this set of totals provides two degrees of freedom of the NP interaction. The remaining two degrees of freedom of the NP interaction are derived from the comparisons among the $\mathcal{J}$ totals where

$$\mathcal{J}_1 = n_0p_0 + n_1p_2 + n_2p_1 \qquad \mathcal{J}_2 = n_0p_1 + n_1p_0 + n_2p_2$$
$$\mathcal{J}_3 = n_0p_2 + n_1p_1 + n_2p_0$$

In the first block in Table I the treatments fall into two groups

$n_0p_0k_0d_0$		$n_0p_1k_1d_0$
$n_1p_1k_0d_0$	and	$n_1p_2k_1d_0$
$n_2p_2k_0d_0$		$n_2p_0k_1d_0$

which have been denoted by $I_1k_0d_0$ and $I_3k_1d_0$ respectively in the margin, and similar considerations apply to the other blocks. Thus it can be seen from the field plan that the design partially confounds $NP(I)$, $NPK(I)$, $NPD(I)$ and $NPKD(I)$.

TABLE I

*Design, field plan, and yields of tubers in cwt per acre (plot size: $\frac{1}{100}$ acre)**

	2200	2010	1211	0001		
$I_1k_0d_0$	175	134	233	154		$I_1k_0d_1$
	0110	1100	1101	0111		
$I_3k_1d_0$	175	175	214	213		$I_3k_1d_1$
	0000	1210	2011	2201		
915	104	152	215	189	1218	
	0211	1201	2211	1111		
$I_3k_0d_1$	266	145	215	188		$I_2k_0d_1$
	2111	0101	0011	1001		
$I_2k_1d_1$	151	184	243	165		$I_1k_1d_1$
	1011	2001	0201	2101		
1067	175	146	174	214	1199	
	1200	0100	1110	0010		
$I_3k_0d_0$	67	105	146	118		$I_2k_0d_0$
	2000	1010	0200	1000		
$I_2k_1d_0$	84	136	112	173		$I_1k_1d_0$
	2110	0210	2210	2100		
750	158	200	174	185	908	

* The underlined figures in the margins are block totals.

The analysis of variance took the form shown in Table III, the error term, with 12 degrees of freedom, consisting of $NPK(J)$, $NPD(J)$, $NPKD(J)$, NKD, PKD, and the unconfounded remainder of $NPD(I)$ and $NPKD(I)$. The responses to, and the two-factor interactions between, nitrogen, phosphate, potash, and dung from this series of experiments have been presented by Reith and Inkson.[3]

Recent interest in some components of these high-order interactions was stimulated by the results of Boyd.[2] With the potato crop he found large interactions between the fertiliser nutrients in the absence of dung, whereas in the presence of dung these interactions were small. For example, dung increased the response to N applied alone, but decreased the response to N applied along with basal P and K. Consequently all experiments of the series were re-examined in order to obtain the linear components of the three- and four-factor interactions. A method described by Binet *et al.*[1] was used to provide estimates of partially confounded interactions freed from block effects.

Analysis

In the first method of analysis the sums of squares for the main effects and first-order interactions were computed from the marginal totals

and the totals in the body of two-way tables. The sum of squares for $NP(\mathcal{J})$ was obtained from the ordinary $\mathcal{J}$ diagonal totals from the two-way $n \times p$ table of $(k_0+k_1)(d_0+d_1)$. Each item in the $n \times p$ table has been summed over all four combinations of k and d, viz. k_0d_0, k_0d_1, k_1d_0 and k_1d_1 since the ordinary rules of algebra apply in expanding $(k_0+k_1)(d_0+d_1) = k_0d_0+k_0d_1+k_1d_0+k_1d_1$. The $NP(I)$ and $NPK(I)$ sums of squares are obtained from the two-way $n \times p$ tables of $(k_0+k_1)(d_0+d_1)$ and of $(k_1-k_0)(d_0+d_1)$ respectively. From the sum of these, $NP(I)$ plus $NPK(I)$, is subtracted the sum of squares for the two block contrasts associated with $NP(I)$ and $NPK(I)$. Hence the sum of squares with 2 degrees of freedom for $NP(I)$ plus $NPK(I)$ unconfounded was obtained. Table III contains an example of the analysis of variance from one experiment in the series.

The second method of analysis is based on an extension of Yates's[4] technique of sums and differences for the 2^3 factorial design. The method will be illustrated on one experiment of the series.

Column 1 of Table II gives the treatment combinations in standard order and column 2 gives the corresponding plot yields of potatoes in cwt per acre. These are divided into successive sets of two as shown. Column 3 is then formed by first adding successive pairs of yields down column 2 and then subtracting successively the first member of each pair from the second: e.g. $258 = 104 + 154, \ldots, 389 = 174 + 215$; $50 = 154 - 104, \ldots, 41 = 215 - 174$. This whole operation is repeated on column 3 to form column 4 since there are two factors at two levels. The entries in column 4 are now divided into successive sets of three. Column 5 consists of three groups of items derived from column 4. The first group of twelve items is obtained by adding the consecutive sets of three in column 4: e.g. $2048 = 619 + 677 + 752$. The second group of twelve items in column 5 is formed by subtracting the first item from the third in each set of three in column 4: e.g. $133 = 752 - 619$. The last group of twelve is formed by taking the sum of the first and third minus twice the second for each set of three in column 4: e.g. $17 = 619 + 752 - 2(677)$. Column 5 is then divided into successive sets of three and column 6 is derived from it by a complete repetition of the process which produced column 5.

Two applications of each of the two processes complete this stage of the calculations since there are two factors at two levels and two factors at three levels.

The first item in column 6, 6057, is the sum of the 36 yields in column 2. The other items correspond to the main effects and interactions, each main effect and interaction total appearing opposite the corresponding column 1 numbers which are suffixes for the small letters to denote treatment combinations and suffixes for the capital letters to denote treatment effects. The suffix 1 corresponds to linear and the suffix 2 to quadratic effects. If the suffix is 0 the factor does not appear in the treatment effect; e.g. the treatment effect corresponding to 0100 in column 1 is the linear effect of phosphate denoted by P_1. A treat-

TABLE II

Calculation of treatment effects and confounding pattern for a $3^2 \times 2^2$ experiment

1	2	3	4	5	6	7	8	9
Treatment *n p k d*	Yield (cwt per acre)				Effect Total	Divisor	Block Constant	Block Contrast
0 0 0 0	104	258	619	2048	6057	36	e_0	
0 0 0 1	154	361	677	1969	911	36	e_3	$36E_1$
0 0 1 0	118	289	752	2040	527	36	e_1	
0 0 1 1	243	388	649	420	101	36	e_4	
0 1 0 0	105	286	723	271	255	24	e_2	
0 1 0 1	184	466	597	220	−7	24	e_5	
0 1 1 0	175	338	579	382	183	24	e_0	
0 1 1 1	213	311	708	91	−107	24	e_3	
0 2 0 0	112	389	753	54	−267	72	e_1	
0 2 0 1	174	334	175	38	251	72	e_4	
0 2 1 0	200	212	117	53	665	72	e_2	
0 2 1 1	266	385	128	10	323	72	e_5	
1 0 0 0	173	230	31	133	−8	24	e_1	
1 0 0 1	165	349	81	−52	−200	24	e_4	
1 0 1 0	136	399	159	174	−328	24	e_2	
1 0 1 1	175	309	143	−47	−28	24	e_5	
1 1 0 0	175	364	22	128	41	16	e_0	$12E_3$
1 1 0 1	214	389	55	−88	−41	16	e_3	$12E_5$
1 1 1 0	146	50	103	77	−171	16	e_1	$12E_2$
1 1 1 1	188	125	99	200	79	16	e_4	$12E_4$
1 2 0 0	67	79	180	−94	−101	48	e_2	$-12E_2$
1 2 0 1	145	38	−27	−71	85	48	e_5	$-12E_4$
1 2 1 0	152	62	−55	−44	239	48	e_0	$36E_3$
1 2 1 1	233	66	173	8	−43	48	e_3	$36E_5$
2 0 0 0	84	−8	119	17	150	72	e_2	
2 0 0 1	146	39	−90	−200	98	72	e_5	
2 0 1 0	134	39	25	−84	254	72	e_0	
2 0 1 1	215	42	75	69	−58	72	e_3	
2 1 0 0	185	78	−41	28	411	48	e_1	$12E_2$
2 1 0 1	214	81	4	154	−391	48	e_4	$12E_4$
2 1 1 0	158	62	47	85	−417	48	e_2	$-36E_3$
2 1 1 1	151	81	3	256	25	48	e_5	$-36E_5$
2 2 0 0	175	29	3	324	333	144	e_0	$36E_3$
2 2 0 1	189	−7	19	161	167	144	e_3	$36E_5$
2 2 1 0	174	14	−36	44	−103	144	e_1	$36E_2$
2 2 1 1	215	41	27	118	191	144	e_4	$36E_4$

ment contrast is given by the number in column 6 divided by the appropriate divisor shown in column 7. For example, 255, the item in column 6 opposite 0100, is the total for the linear effect of phosphate, P_1, and thus $P_1 = 255/24 = 10{\cdot}6$ cwt per acre fresh tubers per 40 lb per acre P_2O_5.

Since all the items in column 6 are linear functions of the yields in column 2, a divisor in column 7 is the sum of squares of the coefficients of the yields contributing to the particular treatment contrast. It can be confirmed that the divisors are given by the formula

$$\text{Divisor} = 2^{2+a-b}\, 3^{2-c}$$

where a = order of contrast, e.g. for two-factor interaction $a = 2$
b = number of two-level factors in contrast
c = number of linear terms of three-level factors in contrast.

The sum of squares for a treatment contrast is calculated by squaring the item in column 6 and dividing by the divisor in column 7: e.g. the sum of squares for P_1 is $(255)^2/24 = 2709{\cdot}38$. The sum of these quantities for all 35 contrasts is equal to the sum of squares of the yields in column 2 minus the correction term, $(6057)^2/36$.

We have yet to take account of the block effects and shall denote block constants by e_0, e_1, e_2, e_3, e_4, and e_5. These are subject to the restriction that $e_0 + e_1 + e_2 + e_3 + e_4 + e_5 = 0$ and we define the following block contrasts:

$$\begin{array}{rrrrrrcl}
-e_0 & -e_1 & -e_2 & +e_3 & +e_4 & +e_5 & = & 6E_1 = 6D \\
-e_0 & +e_1 & & -e_3 & +e_4 & & = & 4E_2 \\
+e_0 & +e_1 & -2e_2 & +e_3 & +e_4 & -2e_5 & = & 12E_3 \\
+e_0 & -e_1 & & -e_3 & +e_4 & & = & 4E_4 \\
-e_0 & -e_1 & +2e_2 & +e_3 & +e_4 & -2e_5 & = & 12E_5
\end{array} \qquad \ldots . \ (1)$$

If we replace the yields of column 2 by the appropriate block constants of column 8 and carry out on them the same operations which produced column 6 from column 2, we obtain, with the aid of equations 1, the block contrasts in column 9. (N.B. The numbering of the blocks is not serial with regard to the layout in Table I; it is related to the pattern of allocation of treatments to blocks.) An entry in this column indicates that the corresponding treatment contrast is confounded with the particular block contrast. We find, in fact, that D is confounded with E_1; that N_1P_2, N_2P_1, N_1P_1K, and N_2P_2K are confounded with E_2; that N_1P_1, N_2P_2, N_1P_2K, and N_2P_1K are confounded with E_3; that N_1P_2D, N_2P_1D, N_1P_1KD, and N_2P_2KD are confounded with E_4; and that N_1P_1D, N_2P_2D, N_1P_2KD, and N_2P_1KD are confounded with E_5.

Of those interactions which are partially confounded we are interested in the four components of NP and in the linear components only of the three- and four-factor interactions. The other partially con-

founded three- and four-factor contrasts involving quadratic components of N and/or P are assumed to be negligible. We can then obtain N_1P_1, N_1P_2, N_2P_1, N_2P_2, N_1P_1K, N_1P_1D, and N_1P_1KD adjusted for block contrasts by means of four sets of least-squares equations.

$$\begin{aligned}
24E_2 &- 12N_1P_2 + 12N_2P_1 + 12N_1P_1K = -26\\
-12E_2 &+ 48N_1P_2 = -101\\
12E_2 &+ 48N_2P_1 = 411\\
12E_2 &+ 16N_1P_1K = -171
\end{aligned} \quad \ldots\ (2)$$

$$\begin{aligned}
72E_3 &+ 12N_1P_1 + 36N_2P_2 = 606\\
12E_3 &+ 16N_1P_1 = 41\\
36E_3 &+ 144N_2P_2 = 333
\end{aligned} \quad \ldots\ (3)$$

$$\begin{aligned}
24E_4 &+ 12N_1P_1KD = -12\\
12E_4 &+ 16N_1P_1KD = 79
\end{aligned} \quad \ldots\ (4)$$

$$\begin{aligned}
72E_5 &+ 12N_1P_1D = -40\\
12E_5 &+ 16N_1P_1D = -41
\end{aligned} \quad \ldots\ (5)$$

The coefficients for the treatment contrasts in the leading diagonals of these equations are obtained from column 7 of Table II, and the non-diagonal coefficients are obtained from column 9. The coefficient for the block contrast in the first equation of each set is six times the coefficient of the appropriate E in the defining equations (1). The block contrast sums, -26, 606, -12, and -40 are computed by equations (1) with each e_i replaced by the total of the six plots in block i. For example, the totals for blocks 0, 1, 3, and 4, shown in the margins of Table I, are identified by using columns 1 and 8 of Table II as 915, 908, 1218, and 1199 respectively. Thus, using the second of the defining equations (1), $-915+908-1218+1199=-26$. The other constants in the least-squares equations are obtained from column 6 of Table II for the appropriate treatment contrasts.

The solution of these equations provides combined sums of squares for the treatment contrasts adjusted for block contrasts. For example, from equations (2),

$$\begin{aligned}
E_2 &= -2{\cdot}861\ 111\\
N_1P_2 &= -2{\cdot}819\ 444\\
N_2P_1 &= 9{\cdot}277\ 778\\
\text{and}\quad N_1P_1K &= -8{\cdot}541\ 667
\end{aligned}$$

The combined sum of squares due to these four contrasts is

$$(-2{\cdot}861\ 111)(-26)+(-2{\cdot}819\ 444)(-101)+(9{\cdot}277\ 778)(411)+(-8{\cdot}541\ 667)(-171) = 5632{\cdot}94$$

By subtracting from this the mean square for the block contrast E_2, $(-26)^2/24=28{\cdot}17$, we obtain 5604·77, the combined sum of squares for the three treatment contrasts, adjusted for E_2. From the other sets

of least-squares equations, (3), (4), and (5), we find the sums of squares for N_1P_1 and N_2P_2 adjusted for E_3 (257·29), for N_1P_1KD adjusted for E_4 (722·50), and for N_1P_1D adjusted for E_5 (84·20). The analysis of variance can now be completed and is presented in Table III along with that for the first method.

TABLE III

Analysis of variance for a $3^2 \times 2^2$ experiment

First method

Source of Variation	D.F.	S.S.	M.S.
Blocks (unadjusted)			
D	1	23 053·36	
$NP(I)$, $NPK(I)$	2	5 128·67	
$NPD(I)$, $NPKD(I)$	2	28·22	
N	2	315·17	
P	2	3 699·50	
K	1	7 714·69	
$NP(I)$	2	1 154·17	
NK	2	5 378·72	
PK	2	7 537·39	
ND	2	1 800·05	
PD	2	877·06	
KD	1	283·36	
$NP(I)$, $NPK(I)$ (unconfounded)	2	4 124·23	
Error	12	7 220·16	601·68
Total	35	68 314·75	

Second method

Source of Variation		D.F.	S.S.	M.S.
Blocks (unadjusted)				
$E_1 = D$		1	23 053·36	
E_2		1	28·17	
E_3		1	5 100·50	
E_4		1	6·00	
E_5		1	22·22	
N_1		1	2·67	
N_2		1	312·50	
P_1		1	2 709·38	
P_2		1	990·12	
K		1	7 714·69	
N_1K		1	4 482·67	
N_2K		1	896·06	
P_1K		1	1 395·38	
P_2K		1	6 142·01	
N_1D		1	1 666·67	
N_2D		1	133·39	
P_1D		1	2·04	
P_2D		1	875·01	
KD		1	283·36	
N_1KD		1	32·67	
P_1KD		1	477·04	
N_1P_2, N_2P_1, N_1P_1K	adj. for E_2	3	5 604·77	
N_1P_1, N_2P_2	adj. for E_3	2	257·29	
N_1P_1KD	adj. for E_4	1	722·50	
N_1P_1D	adj. for E_5	1	84·20	
Error		7	5 320·08	760·01
Total		35	68 314·75	

The significance of the treatment contrasts P_1KD, N_1KD, N_1P_1D, and N_1P_1KD can now be assessed directly in the second method of analysis by means of variance-ratio tests in the same manner as the unconfounded main effects and first-order interactions. In this example these four contrasts are all small and far from significant. By inverting the matrices of coefficients from equations (2) and (3) standard errors can be computed for the first five adjusted treatment contrasts given in Table IV. The variance of such a contrast is the product of the corresponding diagonal element of the inverse matrix with the error mean square from the analysis of variance. For example, from equations (3) the matrix of coefficients

$$\begin{bmatrix} 72 & 12 & 36 \\ 12 & 16 & 0 \\ 36 & 0 & 144 \end{bmatrix}$$

has for its inverse the matrix

$$\begin{bmatrix} 1/54 & -1/72 & -1/216 \\ -1/72 & 7/96 & 1/288 \\ -1/216 & 1/288 & 7/864 \end{bmatrix}$$

Thus the variance of N_1P_1 is 7/96 times the error mean square. When a treatment contrast is not confounded its variance is obtained by dividing the error mean square by the appropriate divisor from column 7 of Table II. If N_1P_1 had not been confounded this would give 1/16 times the error mean square for its variance.

Table IV also gives figures for the information on the partially confounded contrasts relative to that of a scheme where they are not confounded and the error variance is the same. For example, for N_1P_1 this is the ratio of the reciprocals of the two variances $(96/7) \div 16 = 6/7$. If all treatment mean squares are replaced by the error mean square we obtain a within-block sum of squares with 30 degrees of freedom and add to this the sum of squares for blocks, not including D, with 4 degrees of freedom. This provides an error mean square of 822·27 based on 34 degrees of freedom for blocks of 18 plots confounding only D. The ratio of this mean square to the error mean square with 7 degrees of freedom is the relative efficiency of the two designs (108·2%). Thus the relative gain in efficiency due to the partial confounding compared with confounding only D is 8·2%. This was not sufficiently great to give any advantage to the partially confounded design as far as the contrasts in Table IV are concerned, since the actual relative information is less than 1 in all seven cases.

All of these adjusted contrasts fail to reach the 5% level of significance. Standard errors can also be computed for N_1P_1KD and N_1P_1D by inverting the matrices of coefficients from equations (4) and (5). This gives an alternative to the variance-ratio tests and where the results of a series of experiments are to be combined, these estimates adjusted for partial confounding and their standard errors will be required.

TABLE IV

Adjusted values and relative information for partially confounded contrasts

Contrast	Adjusted Value	Standard Error (±)	Relative Information	
			Same Variance	Actual
N_1P_1	−4·27	7·44	6/7	0·93
N_1P_2	−2·82	4·59	3/4	0·81
N_2P_1	9·28	4·59	3/4	0·81
N_2P_2	0·03	2·48	6/7	0·93
N_1P_1K	−8·54	9·75	1/2	0·54
N_1P_1D	−2.45	7·37	7/8	0·95
N_1P_1KD	8·50	8·72	5/8	0·68

Discussion

An outline has been presented for a method of obtaining estimates of confounded treatment contrasts together with their standard errors in a $3^2 \times 2^2$ factorial experiment in blocks of six plots. Of the eight experiments in the series, three have a smaller error mean square by method 2 while the greatest increase is one of 60%. The four linear three-factor interactions, two obtained directly and two adjusted for block effects, the four components of the *NP* interaction, and the linear four-factor interaction adjusted for block effects produced only three results which were significant at the 5% level from all eight experiments—no more than might be expected by chance.

The same methods of analysis were applied to a series of single replicate $3^3 \times 2$ factorial experiments in six blocks with three levels each of *n*, *p*, and *k* and two levels of *d*. The three-factor interaction *NPK* has eight degrees of freedom, which can be divided into four orthogonal pairs. Each pair represents the contrasts between the sums of three sets of nine *npk* combinations. The four groups of three sets are denoted by the letters *W*, *X*, *Y*, and *Z* by Yates.[4] The twenty-seven treatment combinations of the three factors at three levels are arranged in sets of nine in such a way that every combination of each pair of factors occurs in each set of nine. This ensures that the main effects and two-factor interactions will be unconfounded if these sets are arranged in blocks. *D* and the *W*, *X*, *Y*, or *Z* components of *NPK* were confounded with blocks. The other main effects, all components of two factor interactions, and linear three-factor components N_1P_1D, N_1K_1D, and P_1K_1D were obtained as unconfounded contrasts, while $N_1P_1K_1$ and $N_1P_1K_1D$ were adjusted for the block contrasts with which they were confounded. All other components of the three- and four-factor interactions were assumed to be negligible and thus the estimate of error was based on 19 degrees of freedom. Compared with the previous method of analysis which did not isolate these five interaction components, nine out of twenty experiments in the series have a smaller error mean square by method 2. N_1K_1D was significant at the 5% level once in the twenty experiments, while the other three linear three-factor interactions were each significant twice but with opposite signs in two cases. It was noted that the experiments which showed these

significant results were also among those which showed an increase in precision due to the second method of analysis.

In deciding which of the two methods of analysis to use the prime object is not to select the method which gives the smaller error mean square, having computed both. The ultimate aim is to obtain an unbiased error estimated from those high-order interactions which can be assumed with confidence to be small in comparison with the errors affecting them. If any high-order component in the estimate of error cannot be regarded as negligible the estimate will be inflated in comparison with the true experimental error obtained from differences between replicates of the same treatment combination. This inflated error, however, may be the correct one to use where it is intended that the results should be applicable over a range of values of the interacting factors.

In the two series of experiments there were few cases of significant high-order interactions and the results already reported by Reith and Inkson[3] cannot be considered invalid in consequence of the re-examination by the second method of analysis.

REFERENCES

[1] Binet, F. E., Leslie, R. T., Weiner, S. and Anderson, R. L. (1955). *Analysis of Confounded Factorial Experiments in Single Replications.* North Carolina Agricultural Experiment Station, Tech. Bull. No. 113.

[2] Boyd, D. A. (1959). *J. Agric. Sci.*, **52**, 384.

[3] Reith, J. W. S. and Inkson, R. H. E. (1958). *J. Agric. Sci.*, **51**, 218.

[4] Yates, F. (1937). *The Design and Analysis of Factorial Experiments.* Imperial Bureau of Soil Science, Harpenden, England.

NOTES AND COMMENTS

At a meeting on the 4th July 1961 of representatives of interested bodies there was formed a National Council for Quality and Reliability. The intention is that the Council shall, just as does the National Committee on Non-destructive Testing in its own field, focus the quality control interests, which form a part only of the interests of its member bodies.

Membership is open, not to individuals, but to societies, institutions, and other bodies; the Royal Statistical Society has become one of the founder members with Dr E. S. Pearson as its representative on the Council. The Council will be administered by the British Productivity Council who have named a Secretary and a Treasurer. The British Productivity Council retains its advisory panel on quality control and intends to continue its conference, seminar, and other educational activities in this field.

* * *

Among institutions which organise courses of instruction on statistical quality control and other statistical topics, a prominent part is played by the Institute for Engineering Production at the University of Birmingham, whose eighth course on statistical quality control took place in October 1961. These courses are residential in character, giving time for informal discussion as well as formal lectures, and last two weeks. Readers may be interested in the projected courses for the first half of 1962: January 8th to 19th, Basic Statistics for Industry; January 22nd-February 2nd, Sales Forecasting and Market Research; February 4th-16th, Production Planning and Control; February 19th-March 2nd, Methodology of Operational Research; March 5th-16th, Statistical Quality Control; March 19th-30th, Design of Industrial Experiments; April 2nd-13th, Work Study; April 30th-May 11th, Operational Research; May 28th-June 8th, Mathematical Foundations of Operational Research. For further information write to The Director, The Institute for Engineering Production, 16 Norfolk Road, Birmingham, 15.

* * *

A rise in membership to over 2200 is shown in the fourth annual report of The British Computer Society, presented to members at their annual general meeting. Events of the year included the licence of the Board of Trade to dispense with the word 'Limited' in the Society's name; new work started on data transmission and character recognition; a move to co-ordinate activities in evolving standards for the computer field; and increased activity internationally. The Society's third conference will be held in Cardiff in September 1962.

* * *

A joint conference, organised by the Research and Industrial Applications Sections of the Royal Statistical Society, was held in Churchill Hall, Bristol University, from Saturday to Monday, 9th-11th September 1961. It was

attended by about 170 members. The formal activities of the conference were organised in a single stream of seven sessions, the subjects and speakers being as follows: Sampling Inspection, Professor G. A. Barnard; Queues, Dr J. Keilson; Renewal Theory, Professor D. R. Cox; Some Problems in the Foundations of Statistics, Professor A. Birnbaum; Spectral Analysis, Professor H. E. Daniels; Operational Research, Mr G. W. Sears; and Error-correcting Codes and Redundant Mechanisms, Dr K. D. Tocher. At each session the discussion was opened by a prepared contribution. There was time between sessions for informal discussions and on Sunday morning and late afternoon many members were able to see something of Bristol and Clifton. The conference was held in beautiful surroundings and was a happy occasion. The thanks of the participants are due to those who organised it.

MEETINGS OF SECTIONS AND LOCAL GROUPS OF THE ROYAL STATISTICAL SOCIETY

GENERAL APPLICATIONS SECTION

10th January 1961. Problems of Measuring Private Direct Investment Abroad for Balance of Payments Purposes: M. M. T. Forde

14th February 1961. The Effects of Television upon Family Life: W. A. Belson

14th March 1961. Some Experiments in the Statistical Analysis of Advertising Campaigns: B. Benjamin

11th April 1961. Statistics of the Pharmaceutical Service: Rita J. Maurice

9th May 1961. Supply and Demand in the Cinema Industry: J. Spraos

INDUSTRIAL APPLICATIONS SECTION

13th January 1961. How Long will a Whale Keep?: H. E. Daniels

7th April 1961. A Method of Calculating the Limits of Dimensional Variation: S. J. Morrison

5th May 1961. Recent Experience in Simulation Experiments: K. D. Tocher

MEDICAL SECTION

31st January 1961. The Demand for Hospital Beds: D. J. Newell

28th February 1961. Problems in the Measurement of the Maturity of the Growing Skeleton: R. M. Acheson

28th March 1961. Cousin Marriages, Random Mating, and the Size of Isolates: J. Hajnal

25th April 1961. Present Trends in International Health Statistics: Y. Biraud

30th May 1961. Out-patient Waiting Times: C. E. Rossiter

BIRMINGHAM AND DISTRICT GROUP

3rd January 1961. Statistical Problems in Organising a Direct Selling Service, or how to Misapply Queueing Theory: A. M. Lee

6th February 1961. Some Problems in Road Accident Research: J. C. Tanner

23rd February 1961. Research Methods in the Social Sciences: W. A. Belson

23rd March 1961. Time Series and Prediction: M. S. Bartlett

25th April 1961. Closed Queues: F. Benson

BRISTOL GROUP

19th January 1961. Applications of Statistics in Accountancy: P. N. Wallis

16th February 1961. How to Read Journals Painlessly: G. M. Clarke

16th March 1961. A Survey of Doctors' and Dentists' Remuneration: W. F. F. Kemsley

EAST MIDLANDS GROUP

9th February 1961. 'Comprehensive Cover'—two films lent by A. E. Reed Ltd. Questions answered by R. Douglas

EDINBURGH GROUP

25th January 1961. Selection: D. J. Finney
9th February 1961. A Problem in Linear Programming: W. E. Duckworth (Joint meeting with the Glasgow Group)
27th March 1961. Statistical Aspects of the 1961 Population Census: J. Travers

GLASGOW GROUP

19th January 1961. Construction of Optimum Experimental Designs: J. Aitchison
9th February 1961. Joint meeting with the Edinburgh Group
14th March 1961. Applications of Statistics to Traffic Engineering: R. Hodgen
19th April 1961. Visit to the factory of Fibreglass Ltd., Glasgow

MERSEYSIDE GROUP

10th January 1961. A Method of Calculating the Limits of Dimensional Variation: S. J. Morrison
14th February 1961. The Indolent Plotter: D. Goldberg
10th March 1961. Applications of Statistics: J. P. L. Truesdale, C. F. A. Roberts, W. J. Jennett, and R. L. Plackett (Joint meeting with the Manchester Section of the Oil and Colour Chemists Association)
11th April 1961. Dynamic Programming: D. V. Lindley

NORTH-EASTERN GROUP

11th January 1961. The Sampling of Road Traffic: F. Garwood
8th February 1961. The place of Factor Analysis in Experimental Psychology: H. J. Eysenck (Joint meeting with the North-East Branch of the British Psychological Society)
8th March 1961. Application of Net-work Methods to a Problem in Linear Programming almost of the Transport Type: Alison Doig
29th March 1961. Machine Interference: F. Benson

SHEFFIELD GROUP

26th January 1961. How Can Statistics Help the Manager?: P. J. Stanley
23rd February 1961. Reconciliation of Degrees of Belief with the Frequency Theory: D. F. Kerridge
30th March 1961. The Estimation of Power Spectra: M. J. R. Healy
27th April 1961. Rapid Simulation Techniques for Queueing Problems: K. D. Tocher

SOUTH WALES GROUP

4th January 1961. New Uses of Computers in Statistics: G. A. Barnard
23rd February 1961. How Long will a Whale Keep?: H. E. Daniels
16th March 1961. Dynamic Programming: D. V. Lindley

BOOK REVIEWS

Sven Danø: *Linear Programming in Industry—Theory and Applications: An Introduction.* Vienna: Springer-Verlag, 1960. Pp. viii+120. $4.75.

The author is Assistant Professor of Economics in the University of Copenhagen, and the book is based on lectures given in various institutions in Scandinavia and the United States. Part of the material has already appeared as articles in a number of journals. Although the book is written in English it has a strong American flavour, both in the spelling and in the examples quoted.

It is intended to provide a university textbook on linear programming for students of economics or operational research interested in the theory of production and cost and its practical applications. Secondly, it is the author's hope that engineers, business executives, managers, and others responsible for the organisation and planning of industrial operations may find the book useful. Hence it gives a good number of examples from industrial practice and does not overdo the purely theoretical aspects.

After a brief introduction, the author deals with the underlying principles and the simplex method by means of simple algebraic and graphical techniques. In the following chapters theory and practice are nicely balanced and the applications to blending problems, machine utilisation, inventory problems, and transportation problems are described, with practical examples. So far the methods of solution used have been related to the underlying theory, for obvious reasons. A further chapter demonstrates the actual computational procedures and standard tabular forms used in the simplex and transportation methods, and indicates short-cuts. Short chapters discuss other aspects, such as duality, and an Appendix includes certain proofs.

I wish that the number of footnotes had been drastically reduced: there are sometimes as many as five on a page. It is a matter of personal taste, but I find it easy to miss (or otherwise ignore) the numbers indicating the footnotes, and then arrive at the foot of the page to encounter a mass of material in small print which has then to be related to the appropriate part of the text; this irritates me. If the notes are worth putting in at all, are they not worth putting in parentheses in most cases?

Despite this small criticism I would recommend this book. At times the businessman may find it heavy going, but not unduly so, considering the subject. Although only linear relationships are involved, the explanations of problems with many variables and equations are bound to make demands on the reader. I may say that a colleague with practically no previous knowledge of linear programming found it helpful. It is a pity that the price is high for a book of this size.

D. G. BEECH

Robert Loveday: *A Second Course in Statistics.* Cambridge University Press, 1961. Pp. xi+155. 10s.

Mr Loveday's *A First Course in Statistics* was reviewed in *Applied Statistics*, **8**, 136. In his new volume he explains that the keyword of that first book was observation, and that the important idea of the new one is probability. Of the thirteen chapters the six introducing the ideas of probability, of

sampling (his sampling bottle should be useful in practical class work), of significance, of the χ^2 distribution, and of quality control (with charts for mean, range, and fraction defective) are good. There are also chapters on the normal, binomial, and Poisson distributions, on the method of least squares, and on correlation (in the latter, sections 120-123 and 136-139 repeat sections in the earlier book: some other tables and graphs from that appear here also). In addition to these chapters there are answers to some 170 questions, a Glossary, an Index, and an Appendix of six statistical tables, with other tables, as in the first book, from the *Cambridge Four-Figure Mathematical Tables*. I cannot find his Tables A1 and A6 in my copies of the sources to which he refers (note: there is an index misprinted in the formula below his Table A4). He also gives reference statistical tables on pp. 82, 86, and 109: the last does not agree with the corresponding table in *Fisher and Yates*.

Mr Loveday suggests that this new book will meet the needs of students preparing for statistics as an advanced or scholarship level subject in the General Certificate of Education, and he quotes from questions set by the London, Northern Joint, and Cambridge Boards (as well as the Royal Statistical Society and Association of Incorporated Statisticians): it would not, however, be suitable for such as the Durham scholarship level G.C.E.

In a number of matters I find that the book is confusing, confused, or unsound. Such are: the two estimates of a on p. 49; the two values of the mean on p. 65; the two values in the last lines of pp. 95 and 97; the frequency table of p. 116, question 1; the definition of contingency table; the idea of null hypothesis; the relation between fiducial and confidence limits; the definition of Bernoulli's Theorem; and the occasions when $(n-1)$ should be used and not n in calculating a variance. I am surprised that a teacher should consider it sufficient to give, as his only comment, such a superficial explanation of the difference between the two groups of boys on pp. 76-77.

In general, however, this *Second Course* is a better book than the *First Course*, and it should, on the whole, be useful, with its many examples and exercises (I have not checked his answers), for the persons for whom it is intended. Its production is, of course, up to the usual high level of the Cambridge University Press.

FRANK SANDON

W. D. Furneaux: *The Chosen Few: An Examination of Some Aspects of University Selection in Britain*. Oxford: University Press for the Nuffield Foundation, 1961. Pp. xxvi + 245. 25s.

Immediately after the Second World War there was a great increase in the number of applications for university places—some university departments had twenty applicants for each place. The minimum standards were raised, but even so there were some 20 per cent of those admitted who failed in the end to obtain a degree qualification. In 1948 the Nuffield Student Selection Research Unit was established to conduct a ten-year enquiry into the nature of the qualities desirable in a university student and into the means of detecting these. Mr Furneaux was appointed psychologist in charge, and the present book is a first report. It does not deal with the long-term problems, but with selection as actually effected.

In the course of the study the book deals with various aspects of selection previous to university application. Such are sex, social-economic class (Mr Furneaux seems to confuse 'occupation' and 'industry'), and type of school (he does not seem to be aware of facts such as those recently reported by the

Oxford Magazine,[3] where it was stated that students at Oxford from 'independent' schools form over 80 per cent of the number at Trinity College, and only about 30 per cent at St. Peter's Hall and at St. Catherine's). This is not the place to make many comments on these aspects, though it may be said that perhaps a wider experience of the 11 plus examination (he appears to overlook the range of the ratio—grammar school admissions : age group —of from 8·5 per cent for Rutland to 63·5 per cent for Merioneth, as shown in a return of the Ministry of Education[1]), and of the peculiarities of the north-country towns in which a number of the grammar schools of his sample were situated, might have modified some of his observations.

The investigation here reported is in the main a statistical one. He gives many tables in the book, most of them based either on official returns or on a sample arranged through the good offices of Sheffield University. The immense amount of detail available must have presented considerable difficulty in organising the material for the book, but I am not altogether happy that the solution he has adopted is the best. He found that the official statistics were (as so often) in such form that much difficulty was experienced in combining them, and points out (how many times is this necessary?) that the first essential for any research of this kind is a first-class record system. The resulting statistical pictures thus involve a very considerable amount of extrapolation, assumption, and estimation. He claims that the statistics in his analysis have invariably survived the application of the appropriate tests of statistical significance. I cannot follow his figures at the bottom of p. 23, nor his 'applicants' life table' on p. 34. Neither advanced nor specialist statistical techniques are used, except when he employs Guilford's 'index of forecasting efficiency' (a function of r). The correction usually referred to as that for homogeneity he calls 'correction for range'.

The usual routine for a prospective university student is for him (or her) to start a two-year sixth form course, having obtained a number of Ordinary level (subsequently called O level) passes in his General Certificate of Education (G.C.E.); Furneaux makes no reference to the different types of 'school clientele' of the different G.C.E. examining bodies. After one year there the student submits, in the Autumn term, applications to a number of different universities (or colleges), usually in one specific faculty, for admission in the following October. The Admissions Officer in each university faculty thereupon places him in one of three different categories: (*a*) provisional acceptance (on condition that the applicant obtains the appropriate G.C.E. Advanced level passes,) to cover about 70 per cent of the places available; (*b*) reserve (for vacancies arising from places not filled from the provisional acceptances; or (*c*) rejections (usually on account of O level work not being up to standard). In filling these categories, the Admissions Officer considers the O level results (and any A level if already obtained), reports from the schools, usually, of course, in unstandardised form (it is not stated how the Admissions Officer compares school with school, a problem that arises in other connections also[5]), and, in some cases, interviews. When the A level results are known, in August, the Admissions Officer proceeds to make his offers. On account of multiple applications, many of his offers are not taken up, and the reserve candidates are approached. This means a rushed and complicated job. In 1949, for example, Sheffield had to send out 1041 offers to recruit 522 students. Furneaux estimates that, in the period reviewed by his book, about 80 per cent of the qualified candidates

were finally admitted to a university (the report does not deal with teachers' training or technical colleges).

Furneaux looks at this procedure. He finds that, as elsewhere, the interview seems to be of no practical value in obtaining a fair selection. The information available from the schools, if it could be properly communicated, ought to help more than it does. But a headmaster's letter is of very modest validity: a standardised report form (such as has been, apparently since the book was drafted, prepared by the Committee of Vice-Chancellors and Principals of the United Kingdom together with the Incorporated Association of Head Masters to be somewhat on the lines of the National Institute of Industrial Psychology seven-point scheme) could be rather more useful. The G.C.E. results are easily the best guide (especially if attention is paid to the fact of passing first time), though the tendency has been for the Admissions Officer to use these only in the middle of the range of applicants. The whole procedure is very expensive of the time of the applicant, the headmaster, and the Admissions Officer, and, for interview, of the student's money. And it is not efficient. In medicine, of the men who were accepted by Sheffield 89 per cent later qualified, and of those who were rejected, 45 per cent qualified elsewhere. In faculties other than medicine there was no evidence at all that the applicants who were rejected by Sheffield were in any way academically inferior to those who were accepted elsewhere for an admission in the same year. Some of the applicants rejected by every university obtained external degrees—in technology 78 per cent of the external students qualified, almost the same proportion as the internal ones.

Reviewing all this, Furneaux recommends that applicants be advised to make multiple applications (up to three or four), and that they all be made through a centralised 'clearing house'. He suggests that the G.C.E. A level should be taken earlier in the final school academic year. The later stages of his procedure do not seem as sound, speedy, or simple as those which have been put forward, since the book was drafted, by Professor Pitt.[4] This scheme requires that all relevant data be fed to a computer which is programmed to give decisions based on the preferences of the students and of the Admissions Officers. It should be noted that the decision on the use of five grades only with which to state G.C.E. results, as recommended by the Ministry of Education[2] to eliminate 'markgrubbing' (which it is unlikely to do) will lead to a great deal of bracketing.[6] Pitt's scheme avoids the suggestion of Furneaux that perhaps the universities could leave the ranking entirely to the clearing house.

In some ways the book is, inevitably in these days of rapid change, already out of date. There are no references to comprehensive schools, no references to the installation of a 'clearing-up house' for certain faculties, no reference to the Third Report of the Ad Hoc Committee of the Association of Universities of the British Commonwealth on Procedures for Admission to Universities. The bibliography is rather limited. There are no references to some of the important figures and discussions that have appeared in the educational literature (e.g., in the *Universities Quarterly* or publications of the Northern Universities Joint Matriculation Board. It may be noted that the word *analogous* is wrongly spelt five times.

FRANK SANDON

[1] Ministry of Education (1957). *List 69(1956)*. HMSO, London.
[2] Ministry of Education (1961). *Circular 9/61*. HMSO, London.
[3] *Oxford Magazine*. Quoted in *Times Educational Supplement*, 5th May 1961, p. 904.

[4] PITT, H. R. 'Admission procedure: An experimental system discussed', *Times Educational Supplement*, 26th May 1961, p. 1094.
[5] SANDON, F. (1956). 'A regression control chart for use in personnel selection', *Applied Statistics*, **5**, 20.
[6] SANDON, F. (1959). 'Categories in rating scales: How many and how big', *Brit. J. Educ. Psychol.*, **29**, 155.

PUBLICATIONS RECEIVED

M. S. Bartlett: *An Introduction to Stochastic Processes (Paper edition)*. Cambridge University Press, 1960. Pp. xiv + 213. 22s. 6d.

Cambridge University Press, pursuing their policy of issuing certain text-books in 'paper-back' editions for the benefit of students, have now issued Professor Bartlett's book, which was first published in 1955. This edition is substantially the same as the first edition, a few errors having been corrected; a review has already appeared in this journal (March 1956), so no further review is called for. The cloth edition now costs 42s. The cheaper binding, actually thin card, therefore results in a considerable reduction in price. Although this was intended to appeal chiefly to students, it may also interest those who are prepared to do their own binding.

C. E. Weatherburn: *A First Course in Mathematical Statistics (Paper edition)*. Cambridge University Press, 1961. Pp. xv + 277. 18s. 6d.

The appearance of Professor Weatherburn's book, first published in 1946 and reprinted several times, as a 'paper-back' may be taken as evidence of its continued popularity. It is virtually the same as the first edition, apart from a few minor additions and corrections. This may be accounted a defect by some, on the grounds that fifteen years of progress in statistics demand more extensive revision, yet this can hardly be expected in a cheap edition. Furthermore, the author claims the book to be an exposition of statistical mathematics and may well argue that he sees no reason to depart from his original scheme of presentation. The book covers a fair amount of ground concisely yet without making excessive demands on the reader's mathematical ability.

It is a pity that so little space is devoted to a discussion of the method of maximum likelihood; the reader might well conclude that it is of little importance. Some wry smiles may be induced by the description of limits as fiducial limits, or 'confidence limits'—a calm treatment of a stormy subject!

Nathan Goldfarb: *An Introduction to Longitudinal Statistical Analysis*. Illinois: The Free Press of Glencoe, 1960. Pp. xiii + 220. $5.00.

After briefly recounting the history of statistical method, Professor Goldfarb deals with the nature of the longitudinal method, which he describes as the method of repeated observations from a fixed sample. He compares it with other methods and discusses some of the problems associated with the method. Subsequent chapters deal with such topics as planning and organisation, mechanical processing methods, coding problems, and methods of presentation of results. Finally, some applications, actual and suggested, are described.

THE HOUSEHOLD EXPENDITURE ENQUIRY OF THE MINISTRY OF LABOUR

VARIABILITY IN THE 1953-54 ENQUIRY

W. F. F. KEMSLEY
Government Social Survey, London

The 1953-54 enquiry relied on record books for most of the information on expenditure. These records were kept by each household for a period of three weeks, and this paper examines the variation between the three weeks for a wide range of items. The existence of a significant difference between expenditure in the first week and the other two weeks is demonstrated, and the variation in this difference analysed by type of area and season of the year.

Introduction

The information for the 1953-54 enquiry into Household Expenditure by the Ministry of Labour[1] was obtained partly by interview and partly by written records kept by individual members of each household in the sample. The starting dates of the records were staggered so that the whole period of the enquiry was adequately covered, and an analysis into different periods is included in the Report. For that purpose each record was assigned to a particular period according to the starting date of the record.

For various reasons it was not convenient to analyse the three weeks of the record separately, and this paper now gives such an analysis. The variation between the three weeks is of considerable interest since it has a bearing on the subject of accuracy and of errors in the recorded data. Similar results have been noted on other occasions: by the author,[2] also in a joint paper with J. L. Nicholson,[3] and more recently in a paper by R. Turner.[4] Further analytical work on the expenditure enquiries is still being carried out.

Variation in Expenditure between the Three Weeks

For the tabulations for the main enquiry each item of expenditure had been transferred to a separate punched card, and these primary cards had then been summarised for each household to produce cards showing the total expenditure by a household in each expenditure category for the whole three weeks' period of record keeping. For the purposes of this paper it was necessary to go back to the original primary cards in order to examine the variation between the three weeks. The number of these cards was very great, running into several millions. In order to keep the tabulating for the present investigation within manageable limits, a sub-sample was drawn. The selection was made by taking every sixth household when households

were arranged in order of reference number: that is in order of region, area, and serial number of the household within that area. The characteristics of this sub-sample, consisting of 2147 households, were examined and found to be very close to the corresponding characteristics of the whole sample of 12 911 households.

The average expenditure of the 2147 households in each of the three weeks is given in Table I for the main categories of expenditure; further breakdowns are given in Tables IA and IB. The classification and definition of each item is shown in Appendix C of the Ministry of Labour Report,[1] but certain items do not appear in the present table because information for them was obtained from the preliminary interviews and not from the three-week record books.

TABLE I

Variation of household payments between the three weeks in shillings per household

1953 Code*	Item	Week 1	Week 2	Week 3	All 3 Weeks	Week 1 less ½(Weeks 2+3)
2·1	Cereal foods	11·99	11·46	11·60	35·05	+0·46
2·2	Meat	18·11	17·45	17·34	52·90	+0·71
2·3	Fish	2·33	2·18	2·15	6·66	+0·16
2·4	Oils and fats	4·16	4·13	4·08	12·37	+0·06
2·5	Dairy produce	13·07	12·58	12·47	38·12	+0·55
2·6	Fruit and vegetables	10·68	10·20	10·37	31·25	+0·39
2·7	Sugar and preserves	6·48	6·16	6·19	18·83	+0·31
2·8	Beverages	4·22	3·89	3·93	12·04	+0·31
2·9	Other foods	10·25	10·24	10·67	31·16	−0·21
2 –	All food	81·29	78·29	78·80	238·38	+2·74
3 –	Drink and tobacco	24·06	23·59	23·36	71·01	+0·59
4 –	Clothing and footwear	33·45	27·10	31·11	91·66	+4·35
5 –	Household durables	17·48	16·86	14·23	48·57	+1·93
6 –	Other goods	17·44	16·52	16·36	50·32	+1·00
7 –	Travel	15·30	13·40	12·84	41·54	+2·18
8 –	Services	20·99	19·55	20·29	60·83	+1·07
9 –	Other payments	9·28	8·40	8·31	25·99	+0·92
	Housing and fuel (taken from records)	6·75	5·60	5·74	18·09	+1·08
	All Items	226·04	209·31	211·04	646·39	+15·86

* For a definition of the 1953 code see Appendix C of the Report.[1]

A comparison between the figures in these three tables and the corresponding items in the Report shows slight differences. The majority are within the range likely to arise through sampling fluctuations. Two discrepancies are perhaps worth noting. The first concerns clothing and footwear acquired through clubs. For the Report (see para. 49) adjustments were made, substituting cash payments plus club acquisitions for cash payments plus club payments; these adjustments were made to the summary cards from which the tables in the Report

TABLE Ia

Variations between the three weeks: a detailed analysis of food payments in shillings per household

1953 Code (Last 2 Digits)	Item	Week 1	Week 2	Week 3	All 3 Weeks	Week 1 less ½(Weeks 2+3)
11	Bread, rolls, etc.	4·46	4·33	4·35	13·14	+0·12
12/13	Flour, cereals	2·32	1·98	2·10	6·40	+0·28
14/15	Biscuits, cakes	5·21	5·15	5·15	15·51	+0·06
	All cereal foods	11·99	11·46	11·60	35·05	+0·46
20-22 & 26 23-29 excl. 26	Fresh meat	7·38	7·17	7·03	21·58	+0·28
	Other meat	10·73	10·28	10·31	31·32	+0·43
	All meat	18·11	17·45	17·34	52·90	+0·71
	All fish	2·33	2·18	2·15	6·66	+0·16
	All oils and fats	4·16	4·13	4·08	12·37	+0·06
51	Milk, fresh	6·80	6·55	6·50	19·85	+0·28
53/54	Other milk and cheese	1·75	1·71	1·69	5·15	+0·05
55	Eggs	4·52	4·32	4·28	13·12	+0·22
	All dairy produce	13·07	12·58	12·47	38·12	+0·55
61	Potatoes	2·33	2·22	2·23	6·78	+0·10
62/63 & 69	Vegetables	4·23	4·02	4·10	12·35	+0·17
64/65	Fruit	4·12	3·96	4·04	12·12	+0·12
	All fruit and vegetables	10·68	10·20	10·37	31·25	+0·39
71	Sugar	1·73	1·62	1·69	5·04	+0·08
72	Preserves	1·23	1·08	1·06	3·37	+0·16
73	Sweets, chocolates	3·52	3·46	3·44	10·42	+0·07
	All sugar and preserves	6·48	6·16	6·19	18·83	+0·31
81-83	Tea, coffee, cocoa	3·24	2·99	3·03	9·26	+0·23
84	Soft drinks	0·98	0·90	0·90	2·78	+0·08
	All beverages	4·22	3·89	3·93	12·04	+0·31
91	Icecream	0·50	0·54	0·53	1·57	−0·04
92/99	Other items	2·38	2·23	2·46	7·07	+0·03
93	Meals outside	7·37	7·47	7·68	22·52	−0·20
	All other foods	10·25	10·24	10·67	31·16	−0·21
	All food	81·29	78·29	78·80	238·38	+2·74

were compiled, but not to the primary cards used for the present data. The second concerns payments for fuel; the figures in the Report were based partly on information obtained by interview and partly from expenditure recorded in the books. The figures here relate solely to the latter.

It will be seen from the three tables that there is little or no difference in the expenditure recorded in the second and third weeks. On the other hand, the first week is generally above the second and third. The increased expenditure recorded in the first week over the mean of the other two weeks is shown in the final column of the three tables. The average discrepancy is about 7 per cent, but although the same pattern appears in most categories and items, it varies considerably in magnitude. The discrepancy is particularly marked in purchases of clothing and household durables; in travel most of the variation is attributable to the purchase of motor vehicles. In particular, one item

TABLE Ib

Variations between the three weeks: a detailed analysis of non-food payments in shillings per household

1953 Code	Item	Week 1	Week 2	Week 3	All 3 Weeks	Week 1 less ½(Weeks 2+3)
3·1	Alcoholic drink	7·72	7·75	7·66	23·13	+0·02
3·2	Tobacco	16·34	15·84	15·70	47·88	+0·57
	Drink and tobacco	24·06	23·59	23·36	71·01	+0·59
4·1	Men's clothing†	8·32	5·74	6·95	21·01	+1·98
4·2	Women's clothing†	9·59	8·21	9·39	27·19	+0·79
4·3, 4·4 & 4·51	Children's clothing†	3·26	2·53	3·25	9·04	+0·37
4·5 & 4·6	Other clothing‡	2·76	2·47	3·02	8·25	+0·01
4·71	Clothing clubs§	4·49	4·24	4·34	13·07	+0·20
4·14	Footwear††	5·03	3·91	4·16	13·10	+1·00
	All clothing & footwear	33·45	27·10	31·11	91·66	+4·35
5·1/2	Furniture & floor coverings	7·31	6·75	7·22	21·28	+0·32
5·3	Soft furnishings	3·64	2·30	1·92	7·86	+1·53
5·4/5	Radio, gas/elect. appliances	4·13	5·56	2·91	12·60	−0·10
5·6	China, glass	2·40	2·25	2·18	6·83	+0·18
	Household durables	17·48	16·86	14·23	48·57	+1·93
6·1/36/37	Jewellery, optical goods	1·88	2·32	2·19	6·39	−0·37
6·21/22	Books, newspapers	4·39	3·71	3·72	11·82	+0·67
6·23/24	Toys, stationery	1·65	1·62	1·90	5·17	−0·11
6·31-34	Medical supplies	1·21	1·12	1·20	3·53	+0·05
6·35	Toilet requisites	1·64	1·57	1·61	4·82	+0·05
6·4	Non-durables	3·87	3·51	3·53	10·91	+0·35
6·5	Miscellaneous	2·80	2·67	2·21	7·68	+0·36
	Other goods	17·44	16·52	16·36	50·32	+1·00
7·1	Purchase of motor vehicles	2·14	0·50	0·36	3·00	+1·71
7·2	Maintenance of motor vehicles	4·02	3·30	2·80	10·12	+0·97
7·3/4	Purchase of cycles, prams	0·98	1·08	1·03	3·09	−0·08
7·51/61/71	Travel by rail or tube	1·62	1·81	2·11	5·54	−0·34
7·52/62/72	Travel by bus	5·85	6·11	5·72	17·68	−0·06
7·73/74/81	Other travel	0·69	0·60	0·82	2·11	−0·02
	All travel	15·30	13·40	12·84	41·54	+2·18
8·1	Postage, telephone, telegram	1·19	1·00	0·98	3·17	+0·20
8·21	Cinemas	2·46	2·44	2·36	7·26	+0·06
8·22-29	Other entertainment	2·89	2·82	2·95	8·66	+0·00
8·4	Domestic help	2·32	2·23	2·22	6·77	+0·09
8·5	Personal service	4·96	4·16	4·37	13·49	+0·70
8·64-66	Professional service	1·71	1·34	1·66	4·71	+0·21
8·72	Other services‡‡	5·46	5·56	5·75	16·77	−0·19
	Services	20·99	19·55	20·29	60·83	+1·07
9·21/99	Other expenditure	1·49	1·39	1·40	4·28	+0·09
8·74/9·42	Subscriptions, Xmas clubs	2·82	2·44	2·49	7·75	+0·36
9·41	Savings	1·42	1·28	1·10	3·80	+0·23
8·31-33	Betting	3·55	3·29	3·32	10·16	+0·24
	Other payments	9·28	8·40	8·31	25·99	+0·92
1·64	Materials for repairs and decorations	2·10	1·83	1·71	5·64	+0·33
1·9	Solid fuel	4·65	3·77	4·03	12·45	+0·75
	Housing & fuel taken from records	6·75	5·60	5·74	18·09	+1·08
	All non-food	144·75	131·02	132·24	408·01	+13·12
	Food	81·29	78·29	78·80	238·38	+2·74
	Total	226·04	209·31	211·04	646·39	+15·86

† Excluding footwear. ‡ Excluding 4·51 and 4·65. § See para. 49 of reference 1.
†† 4·14, 4·24, 4·34, 4·44 and 4·65. ‡‡ 8·72, 8·75, 8·76, 8·77 and 1·51.

contributed about 1·4 shillings to the total discrepancy of 2·18 shillings.

The question which immediately comes to mind is how does this variation come about? Is it due wholly or partly to some defect in recording, such as the entry in the first week of items of expenditure made previously, or to a falling off in the conscientiousness with which a record is kept? Can some of the variation be explained by the respondents actually changing their behaviour while keeping the records?

An internal examination of these data cannot hope to answer such questions. What this paper sets out to do is to determine how far the pattern of variation shown in Table I is really significant, and not simply a chance product of sampling fluctuations. In other words, is the expenditure in the first week significantly higher than in the other two weeks? If this is true of total payments, does it also apply to each category where the first week is highest? Further, the second week is very slightly below the third week. This might be held to support the hypothesis that apparent over-spending in the first week was partially off-set in the second by under-spending. Is this difference significant?

Method of Analysis

The analysis in the following sections attempts to answer the questions posed in the concluding paragraph of the preceding section. A direct method of analysis would have been to examine in detail the variation between the three weeks for each of the 2147 households, using either the actual difference between the weeks in pairs, or some function of these differences. Instead, the study reported in this paper is based not on individual households, but on grouped data. The reason for this is two-fold.

As may be inferred from certain remarks above, the data for the three separate weeks for a single household were punched on separate cards. This was convenient for the main tabulations, but not so satisfactory when examining differences between pairs of weeks. It would have been necessary either to re-punch the data, or perform some equivalent operation on a computor. It was not practicable to do this, but even had it been possible, it is doubtful whether the extra computing would have been justified by the results.

This doubt arises because it is unlikely that the difference between expenditures of a household in two different weeks will follow the normal law. Expenditure in any one week usually has a skew distribution with a low frequency of high amounts, and a proportion of households without expenditure. These 'nils' occur in the broadest of expenditure categories, and they tend to increase the finer the category. Although it seems likely that the difference between expenditure in two weeks will follow a uni-modal distribution, there is no *a priori* justification for assuming that it will be normal. On the contrary, its shape will vary with the proportion of 'nils' in the parent distribution. It is possible, of course, that by a careful selection of separate items an

expenditure category can be so defined as to yield a distribution of differences which is normal, but a selection of this nature severely restricts research to a few special categories.

In these circumstances it would have been unwise to embark on an extensive investigation without first looking into the actual forms of the distributions to see how far they are amenable to conventional significance tests, and if not, whether suitable transformations could be found to give normal distributions. Such a study is likely to be very time-consuming and to make considerable calls on computing resources. In this paper the problem has been side-stepped in a simple manner, although it must be admitted that the results are to some extent approximate. None the less, it has been possible to derive from the exercise several conclusions with only a modest use of computing resources.

For this paper the main emphasis has been put on an analysis where the expenditure amounts in the three weeks have been ranked in order of size. With three weeks, there are six possible arrangements, as listed in the next section. If there were no inter-week bias of the kind discussed above, and no association between any two weeks, each could be regarded as a random and independent selection from the same universe. It would be an even chance whether the expenditure amount for one week was above that for any other week or below it. On this hypothesis the six possible patterns of expenditure would occur with equal frequency. In principle, this argument could be applied to individual households were it not for the existence of weeks without expenditure, or more generally, if there were no pairs of weeks with identical expenditure. In a large sample an occasional tie can be tolerated and dealt with, by assigning half the frequency to each of the two patterns associated with the tie, without seriously vitiating the validity of the test; but if ties are frequent the test will break down.

This difficulty can be overcome by grouping households and applying the test to grouped data rather than to data for individual households. The total expenditure for a group of households is unlikely to be identical for any two weeks, and furthermore, the risk of a 'nil' in such a total is very greatly reduced. Beyond a certain point, the size of the group is immaterial since the argument above applies to any distribution whatever its shape. It is true that the way in which households are grouped may be of importance, and some methods of classification may be more informative than others. That adopted here was determined mainly by convenience, and for this reason was based on the main features of the sample design. The 2147 households in the subsample were aggregated into 68 groups divided among the main strata as follows:

	Quarters of 1953-54			
	1st	2nd	3rd	4th
Larger urban areas	10	10	10	10
Smaller urban areas	5	5	5	5
Rural areas	2	2	2	2

Each group of households in each of the strata has been arranged so as to be a random selection and representative of that stratum. Further details, with the number of households in each group, will be found in the Appendix to this article. The analysis in the next section is based on the total expenditures of the households in each of the 68 groups. This forms the main basis of the paper, but since in carrying through the analysis totals of expenditure in each of the 68 groups were computed it seemed useful to extend the exercise and carry out analyses of variance using these totals as the basic variables. The results are given in a later section, but they are subject to certain limitations which are explained in the Appendix. For this reason, that stage of the work has been limited to broad categories of expenditure and is of secondary importance compared with the analysis in the next section.

Patterns of Expenditure

As explained above, the three separate weeks were ranked by size of total expenditure for each of the 68 groups separately, 1 being assigned to the week with the highest expenditure and 3 to the week with the lowest. The six patterns possible are:

1st week	1	1	2	3	2	3
2nd week	2	3	1	1	3	2
3rd week	3	2	3	2	1	1
Pattern type	A	B	C	D	E	F

From the pattern displayed by the figures of average expenditure in Table I it can be expected that patterns of Type A and B will predominate. A simple test of the variations in Table I can then be obtained by examining the consistency of the patterns in the 68 areas, and the numbers of each pattern type are shown in Table II. Following the argument of the previous section, a distribution-free test of the significance of the inter-week variation was applied by comparing the actual distribution in Table II with the expected frequency for each type, namely 11·33.

TABLE II

Pattern of the inter-week variation in total payments in record books

Pattern Type	Number observed
A	22
B	19
C	10
D	7
E	7
F	3
Total	68

The calculated value of χ^2, with 5 degrees of freedom, was 24·83, and, under the null hypothesis of equal frequency of patterns, the probability of obtaining this calculated value is less than 0·0002.

TABLE III

Probability, P, of equalling or exceeding the values of χ^2 obtained from distributions of patterns

Item†	Expenditure Amounts		Numbers of Purchases	
	Modal Type		Modal Type	
	A or B	Other	A or B	Other
Food				
Bread, rolls	0·117		0·334	
Flour, cereals	0·002		0·173	
Biscuits, cakes	0·430			0·057
Fresh meat	0·790		0·894	
Other meat	0·120		0·007	
Fish	0·075		0·214	
Oils and fats		0·476	0·082	
Milk	0·023		0·339	
Other milk and cheese	0·018			0·334
Eggs	0·033		0·020	
Potatoes	0·075		0·050	
Vegetables	0·023		0·148	
Fruit	0·092		0·768	
Sugar	0·682		0·344	
Preserves	0·008		0·010	
Sweets, chocolates	0·871		0·493	
Tea, coffee, cocoa	0·010		***	
Soft drinks	0·135		0·003	
Ice cream		0·010		0·993
Other items		0·208	0·094	
Meals outside		0·424	0·146	
Non-food				
Alcoholic drink		0·628	0·133	
Tobacco	***		***	
Men's clothing	0·020		0·004	
Women's clothing		0·054	0·574	
Children's clothing		0·621	0·800	
Other clothing	0·208		0·062	
Clothing clubs	0·254		0·182	
Footwear	0·176		0·052	
Furniture	0·226		0·481	
Soft furnishings	0·876			0·619
Radio, etc.	0·286		0·570	
China, glass		0·953	0·210	
Jewellery, etc.		0·524		0·198
Books, newspapers	0·233		0·047	
Toys, stationery		0·500	0·334	
Medical supplies	0·075		0·016	
Toilet requisites	0·814		0·036	
Non-durables	***		***	
Miscellaneous		0·208	0·208	
Maintenance of motor vehicles		0·413		0·080
Purchases of cycles, prams		0·887	0·115	
Travel by rail		0·402	0·781	
Travel by bus		0·263	0·045	
Other travel		0·208		0·654
Postage etc.	***		***	
Cinemas		0·735	0·574	
Other entertainment		0·368	0·210	
Domestic help		0·044	0·143	
Personal service	0·002		***	
Professional services		0·329	0·457	
Other services	0·208		***	
Other expenditure		0·446	0·069	
Subscriptions etc.	0·549		0·014	
Savings	0·160		0·051	
Betting	0·006		0·003	
Materials for repairs	0·568		0·125	
Solid fuel	0·084		***	

† For more exact definition of items see Tables IA and IB. *** $<0{\cdot}001$.

The deviation of these figures from the expected values is highly significant, with A and B patterns well above the expected numbers and type F well below. This would seem to establish the predominance of A and B patterns, but the slight excess of A over B is not significant.

Since the sample has been drawn from the general population the basic data are not homogeneous, and it is possible that some parts of the population contributed more to this bias than others; for example, there might be a differential effect with income. This particular hypothesis could not be examined within the limits of the study, but in counting up the various patterns no obvious bias appeared in favour of one or other of the strata. For instance, the numbers of patterns A and B were more frequent than their expected values for each of the three types of areas, and also for each of the four quarters. The fact that there was no differential bias associated with these two characteristics is further confirmed by the analysis of variance reported in a later section.

The preceding analysis has been carried out on each of the individual items shown in Tables IA and IB. The values of χ^2 have been calculated for each individual item in the same way as for Table II; for each χ^2 the probability P of obtaining this or a greater value is given in Table III. Two sets of values are shown: on the left are those for expenditure and on the right the corresponding values when the number of purchases is taken as the variable in the place of total expenditure. Both parts of the table distinguish cases where the modal pattern was type A or B from those where one of the other types predominated.

Considering first the entries where the modal pattern was type A or B, the outstanding feature is the prevalence of low values of P and the scarcity of high values. Not only are there several values of χ^2 above the conventional significance levels of 0·01 and 0·05, but there is a disproportionate number of cases with values just below these levels. This suggests that there is a general tendency for the first week to be highest, both in terms of expenditure amounts and number of purchases. For this reason, categories such as 'furniture' and 'books, newspapers, etc.,' although displaying a value of χ^2 below the conventional level for significance are, nevertheless, probably subject to the same general tendency. Indeed, it is more difficult to say with certainty which items do not display inter-week variation. In many cases, even where the predominant pattern in expenditure was neither A nor B, it is not possible to disregard entirely the suspicion that they are subject to underlying inter-week variation of the same kind. Many items, where the predominant pattern in expenditure was neither A nor B, do however display this bias when the analysis is made by numbers of purchases; few items actually fall into the final column of Table III.

Perhaps only the following items should be considered as not contributing to inter-week variation:

Biscuits, cakes, etc.	Jewellery
Fresh meat	Toys, stationery
Sugar	Maintenance of vehicles
Sweets, chocolates	Travel by rail
Icecream	Other travel
Women's and children's clothing	Cinema
Soft furnishings	Domestic help
China, glass	Professional services

Even so, it is a matter of opinion whether some of the cases listed above should be included. Furthermore, of the items where the modal type is A or B, about one half are A and half B when expenditure is under consideration; whereas when the number of purchases is considered, two-thirds are A and one-third B. There is no support here for the hypothesis that the increased expenditure reported in the first week is balanced by under-spending in the second week as compared with the third.

Finally, there are few items where the predominant pattern is neither A nor B but some other type, and where at the same time χ^2 is above the significance level, and they are therefore best regarded as chance results. In any case, not one of the items in question, namely 'icecream', 'women's clothing', 'domestic help', is consistent as between the expenditure patterns and the corresponding data for number of purchases.

Variability of the Inter-week Variation in Expenditure

The preceding section has demonstrated the existence of a variation between weeks, and has shown that there is a general tendency for expenditure reported in the first week to be above that of the two subsequent weeks. This section continues the analysis by applying the technique of analysis of variance to the totals of expenditure already computed for each of the 68 groups of households. The use of totals based on groups of households as the basic variable has certain drawbacks as well as advantages (see the second part of the Appendix), and for this reason this part of the work is subsidiary to the main analysis and is restricted to broad categories of expenditure.

It is first necessary to explain how total variation in weekly expenditure can be split into a number of components. If each week is treated as a separate value ignoring whether it is the first, second or third, then the variance of the distribution of these weekly values can be divided into two parts. One part is proportional to the sum of the squared deviations of the sample mean for each week from the overall mean, while the other is proportional to $\mathrm{Var}(x_1)+\mathrm{Var}(x_2)+\mathrm{Var}(x_3)$, where Var (x_i) represents the variance of expenditure in week i about the sample mean for that week. Further, if $3\bar{x}=x_1+x_2+x_3$, so that $\bar{x}$ is the expenditure of a household averaged over the three weeks, then

$$\mathrm{Var}(x_1)+\mathrm{Var}(x_2)+\mathrm{Var}(x_3) = \mathrm{Var}(x_1-\bar{x})+\mathrm{Var}(x_2-\bar{x})+\mathrm{Var}(x_3-\bar{x}) + 3\mathrm{Var}(\bar{x}) \quad \ldots . (1)$$

The sum of the first three variances on the right, namely $\mathrm{Var}(x_1-\bar{x})+\mathrm{Var}(x_2-\bar{x})+\mathrm{Var}(x_3-\bar{x})$, may be regarded as the total inter-week variation (excluding bias), and this can be divided into two components using transformed variables y and z where

$$y = x_1 - \tfrac{1}{2}(x_2+x_3) \text{ and } z = x_2 - x_3$$

$$\text{Then } \mathrm{Var}(x_1-\bar{x})+\mathrm{Var}(x_2-\bar{x})+\mathrm{Var}(x_3-\bar{x}) = \tfrac{2}{3}\mathrm{Var}(y)+\tfrac{1}{2}\mathrm{Var}(z) \quad \ldots\ldots (2)$$

This partition may be made in three ways, depending on which week is chosen as being different from the other two. In the light of the preceding discussion, week 1 has been treated as the outstanding one, so that y has been defined as the excess of week 1 over the mean of weeks 2 and 3. The relationship expressed in equations (1) and (2) are identities and apply whether the basic data refer to individual households or groups of households.

The main analysis is concerned with $\mathrm{Var}(y)$ and $\mathrm{Var}(z)$, but the importance of the two relationships (1) and (2) is that they enable the inter-week variation to be viewed in perspective against the variation in the average weekly expenditure. The three values in question are shown in Table IV. Two conclusions follow: the sizes of both components of inter-week variation are similar, the F-ratio of 1·17 being below the level of significance. Thus, if the variation between weeks 2 and 3 is considered to be an estimate of the normal inter-week variation, the variance obtained by comparing week 1 with the two subsequent weeks is of comparable magnitude and not abnormal. Secondly, total inter-week variation (excluding bias) is small compared with the variation in the average weekly expenditure of the three weeks. Both components of inter-week variation are under one-eighth of the variation in average weekly expenditure, and well below the lower 0·1% point of an F-ratio.

TABLE IV

Analysis of the total variation within strata of total expenditure in shillings

Source of Variation	Mean Square*	F-ratios	
$\frac{2}{3}$ Var(y)	273 181	1·17	0·121
$\frac{1}{2}$ Var(z)	233 482	1·00	0·103
Inter-week variation	506 663		
3 Var($\bar{x}$)	2 258 502		1·000
Total	2 765 165		

* Each based on 56 degrees of freedom

Returning to the main subject of this section, namely the two components of inter-week variation $\mathrm{Var}(y)$ and $\mathrm{Var}(z)$, each may be further analysed into:

(i) differences between the three types of area, i.e. large urban; small urban; and rural;
(ii) differences between the four quarters of 1953-54;

(iii) interaction between these two factors;
(iv) variance within the twelve strata.

TABLE V

Analysis of variance of y and z, total expenditure in shillings

Source of Variation	DF	SS		MS		*F*-ratio	
		y	*z*	*y*	*z*	*y*	*z*
(*a*) Between area types	2	809 495	50 328	404 748	25 164	0·96	0·05
(*b*) Between quarters	3	892 838	382 326	297 613	127 442	0·71	0·27
(*c*) Interaction	6	2 072 998	5 211 570	345 500	868 595	0·82	1·87
(*d*) Between strata	11	3 775 331	5 644 224	343 212	513 111	0·82	1·10
(*e*) Within strata	56	23 538 913	26 040 791	420 338	465 014	1·00	1·00
(*f*) Total	67	27 314 244	31 685 015	407 675	472 911		

The appropriate figures for Var(y) and Var(z) separately are shown in Table V. Only one of the mean squares in lines (*a*), (*b*), and (*c*) exceeds the corresponding mean square in line (*e*) of the table; this is the interaction item for z, but the corresponding F-ratio, namely 1·87, is below the 5 per cent level of significance. Thus, it appears that

TABLE VI

Inter-week variation: variability between areas and quarters of expenditure in shillings

Item	*F*-ratios for Variable *y*†			*F*-ratios for Variable *z*†		
	Areas	Quarters	Areas and Quarters	Areas	Quarters	Areas and Quarters
Cereal foods	5·11**	0·33	2·32*	2·14	2·36	1·35
Meat	1·55	0·61	0·94	0·76	0·47	0·65
Fish	0·04	0·54	1·23	0·24	0·40	0·69
Oils & fats	2·19	4·15**	1·87	0·28	0·55	0·60
Dairy produce	0·05	2·82*	1·57	1·41	0·41	0·95
Fruit & vegetables	2·11	1·02	0·99	1·02	0·07	0·83
Sugar & preserves	0·40	0·45	0·64	1·14	0·05	0·47
Beverages	1·09	1·52	1·19	1·44	0·48	1·05
Other foods	0·01	6·11**	1·91*	1·25	0·65	0·71
All food	2·36	2·02	1·22	0·84	0·14	0·44
Drink & tobacco	0·67	0·44	0·74	0·32	2·25	0·77
Clothing & footwear	3·18	0·96	1·98*	2·55	0·15	1·90*
Household durables	0·08	0·42	1·01	0·57	0·76	0·58
Other goods	0·04	0·62	0·40	1·20	0·75	1·07
Travel	3·48*	5·01**	3·16**	5·47**	0·83	2·18*
Services	1·73	0·55	1·67	1·61	0·21	1·53
Other payments	0·38	1·09	1·53	1·22	1·11	1·21
Housing & fuel (taken from records)	2·66	2·33	1·58	0·22	2·25	1·25
All items	0·96	0·71	0·82	0·05	0·27	1·10

† Degrees of freedom as shown in Table V.
Significance: 5% *; 1% **.

inter-week variation is not subject to specific sources of variation arising from differences between the three types of area or between the four quarters of the survey period. However, what is true for expenditure as a whole may not be true for separate categories of expenditure. With this in mind the calculations have been repeated for the same commodity groups as used in Table I. The resulting *F*-ratios are shown in Table VI.

There are few cases of significance; so few, that possibly they may be dismissed as chance events. Even so, it may not be without interest that most of the few ratios that are significant relate either to 'travel' or 'food', and probably are related to seasonal or regional variation.

Approximate standard errors of mean y and z have also been computed for each category. They have been derived from the within-strata variances in line (*e*) of Table V, and the corresponding mean squares used as a basis for Table VI. These standard errors are shown in Table VII beside the corresponding means taken from Table I.

TABLE VII

Standard errors of mean inter-week variation of expenditure in shillings

Items	y: Week 1 less $\frac{1}{2}$(Weeks 2 + 3)		z: Week 2 less Week 3	
	y	S.E.(y)	z	S.E.(z)
Cereal foods	+0·46**	0·10	−0·14	0·12
Meat	+0·71**	0·19	+0·11	0·24
Fish	+0·16**	0·05	+0·03	0·06
Oils & fats	+0·06	0·05	+0·05**	0·01
Dairy produce	+0·55**	0·13	+0·11	0·17
Fruit & vegetables	+0·39**	0·11	−0·17	0·13
Sugar & preserves	+0·31**	0·09	−0·03	0·10
Beverages	+0·31**	0·06	−0·04	0·08
Other foods	−0·21	0·19	−0·43	0·27
All food	+2·74**	0·38	−0·51	0·61
Drink & tobacco	+0·59*	0·32	+0·23	0·45
Clothing & footwear†	+4·15**	1·49	−3·91**	1·54
Household durables	+1·93	2·16	+2·63	2·31
Other goods	+1·00	0·78	+0·16	0·51
Travel‡	+0·47	0·63	+0·42	0·60
Services	+1·07	1·13	−0·74	1·01
Other payments	+0·92**	0·26	+0·09	0·28
Housing & fuel (taken from records)	+1·08**	0·38	−0·14	0·42
All items † ‡	+13·95**	3·13	−1·77	3·29

Significance: 5%*; 1%**.
† Excluding payments into clothing clubs
‡ Excluding purchases of motor vehicles

In preparing the data for Table VII two items of expenditure were discarded: clothing clubs (for the reason explained in para. 49 of the Ministry of Labour Report) and purchases of motor vehicles because, as noted earlier, one item contributed most of the discrepancy. The total excess disregarded was 1·91 shillings, hence the standard error for all items relates to 13·95 shillings per household, and not to the 15·86 shillings shown at the foot of Table I. The significance of the greater amount reported in the first week is beyond question as the

mean excess is nearly 4·5 times its standard error. On the other hand, because the mean squares in lines (*a*), (*b*), and (*c*) of Table V do not differ significantly from the within-strata variance in line (*e*), this excess expenditure is independent of quarter or type of area.

Table VII confirms the significance of the first week bias in many categories. However, it would be unwise to conclude on the evidence of this table alone that this higher expenditure, which also appears in the first week in most other categories, was accidental. The conclusion from the main analysis earlier was that there were few items in which the first week's expenditure was not above that of the other two weeks. It seems likely that even where $\bar{y}$ failed to reach the level for significance, the general tendency still operates in many items. This is borne out by a comparison of Tables III and VII. Categories for which $\bar{y}$ in Table VII fails to reach the level of significance incorporate many of the items which in Table III are inconsistent as between the two methods of analysis. Whereas some of these items did not conform to the general tendency when the distribution of expenditure patterns was examined, nearly all produced a pronounced bias when number of purchases was taken as the variable.

Turning now to $\bar{z}$, Table VII confirms that the slight excess in total payments made in week 3 over week 2 is well below the level of significance. Only two categories show a significant value for $\bar{z}$. The difference for one of these, 'clothing and footwear', is sufficiently large to account for more than the whole discrepancy in total payments, whereas that for the other category, 'oils and fats', is of negligible size. As there are only two significant values in the series, it would seem advisable to disregard the difference between the second and third weeks without further supporting evidence.

Conclusions

This paper examines the variation between expenditure in the three sets of seven days for which spending was recorded by informants in the 1953-54 enquiry. It would appear that:

(*a*) Expenditure reported in the first week differs from that reported in the two subsequent weeks.

(*b*) Subtracting the mean of the second and third weeks from the reported expenditure in the first, produces a positive difference, and this bias is in general independent of factors such as type of area or season of the year.

(*c*) Apart from this bias, the excess expenditure for individual households has a variance which is no greater than that between the second and third weeks.

(*d*) Excluding this bias, total inter-week variation is small compared with the inherent variability of average weekly expenditure.

(*e*) The slight excess in mean expenditure for the third week over the second is not significant and should be disregarded.

These conclusions apply to total expenditure and to many ex-

penditure categories. It remains to emphasise that the analysis described here cannot lead to the underlying causes behind the biases just described. Moreover, whereas the significance of inter-week variation is established, the actual amount of the bias is relative; an internal examination by itself does not make it possible to say whether the second and third weeks are depressed, the first week inflated, or whether the true position lies in between.

APPENDIX

Further Details of the Basic Groups used for the Analysis

For the purposes of analysis the 2147 households in the one-in-six sub-sample were divided into 68 groups, and the expenditure of the households in each group was aggregated and this total was then treated as the basic variable throughout the exercise. In forming these groups the sub-sample was first taken in 12 main strata giving the numbers shown in Table A.

TABLE A

Households in sub-sample classified by type of area and quarter

Quarter:	1st	2nd	3rd	4th	All Quarters
Larger urban areas	212	184	190	210	796
Smaller urban areas	221	207	216	220	864
Rural areas	120	118	121	128	487
All areas	553	509	527	558	2147

Each group of households in the larger urban strata (administrative areas with population in excess of 100 000, including the County of London) was further divided into 10 groups; each smaller urban stratum into 5 groups; and each rural stratum into 2 groups, making 68 in all. Each group was an equivalent sub-sample and representative of the stratum from which it was drawn. In order to explain how this was arranged it is necessary to describe the sample design of the 1953-54 enquiry.

Briefly, in the larger urban strata, there was only one stage of sampling; the households, or more precisely the addresses, were taken from the rating books of local authorities at a fixed interval to provide 7669 addresses. In the other two strata the sampling was carried out in two stages; at the first stage 200 smaller urban areas and 80 rural areas were selected for the whole twelve months covered by the enquiry. The smaller urban areas were chosen from the 1124 such areas in the United Kingdom after the latter had been arranged in alphabetical order within counties; the counties in alphabetical order within regions; and the regions likewise in order. The 200 areas were drawn with probability proportional to population size, and with replacement. A similar procedure was employed in order to select 80 areas from a

total of 544 rural areas. At the second stage of sampling 40 addresses were drawn from each smaller urban area, and 52 from each rural area; as with larger urban areas, the addresses were taken from the rating books of local authorities. Further details of the sample will be found in paras. 11 to 18 of the Ministry of Labour report.[1]

For purposes of the present exercise the 50 areas in each quarter of the smaller urban stratum were further sub-divided into 5 groups, each of 10 areas. This selection was made by drawing a one-in-five interval sample from the primary areas when arranged in the original order of the sample design. The numbers of households in each of the five groups in the four smaller urban strata are shown in the upper part of Table B. A similar procedure was applied to the 20 primary areas in each of the four rural strata to yield two groups each of 10 areas; the numbers of households in these groups are given in the lower part of Table B.

TABLE B

Households in each set of grouped primary units

	Quarters of 1953-54			
	1st	2nd	3rd	4th
Smaller urban areas	42	41	39	46
	44	42	46	41
	46	41	47	48
	43	43	38	42
	46	40	46	43
Rural areas	60	60	58	67
	60	58	63	61

The division of the larger urban stratum presented more of a problem, mainly because there was only a single stage of sampling in this stratum, so that there were no first-stage units which could conveniently form the basis of grouping as in the other strata. A sub-division was effected by dividing the cooperating households in each quarter of the larger urban stratum into 10 sub-samples according to the last digit of the serial number given to the address for identification purposes. These reference numbers were not associated with any particular characteristic of the address, but were allocated according to the order in which the addresses were drawn, starting afresh for each separate administrative area. The numbers of households in each of the 10 groups in each quarter of the stratum are shown in Table C.

In all strata the method of sub-division is approximately equivalent to replacing the original first-stage sampling interval by a wider interval, so that each group is geographically representative of the stratum of which it forms a part. It will be seen that the numbers within a stratum vary from group to group.

Each of the 20 groups in the smaller urban strata would have contained the same number of households, namely about 43, had it not been for variations in response; each of the 10 groups in the rural strata would have contained about 61 households but for the same

TABLE C

Households in the 40 sub-divisions of the larger urban stratum

Quarters of 1953-54			
1st	2nd	3rd	4th
13	15	15	31
29	24	20	26
27	24	23	15
19	21	18	12
20	21	26	22
24	22	22	24
24	16	23	25
13	18	17	18
23	13	12	18
20	10	14	19

cause. The variation in numbers within the larger urban strata is partly due to variations in response and partly due to the fact that the separate areas of which these strata were composed were not all of the same population size, so that the range of serial numbers varied from area to area. This variation was of no consequence so far as the main analysis was concerned, but it had certain implications on the application of variance analysis in the subsidiary part of the investigation.

Variation in Household Numbers and the Analyses of Variance

As remarked early on in the paper, differences between household expenditure in any two weeks are unlikely to follow the normal distribution, although their distribution is likely to be uni-modal in shape. On the other hand, differences obtained from aggregated expenditures for a group of households will tend to follow a normal curve, and to do so more closely the larger the number of households included in the group. Therefore, one of the advantages in using a basic variable derived from aggregating expenditure for a number of households is that it is possible to apply significance tests with greater confidence than would be the case if the tests were employed on data based on individual households. However, this device reduces the degrees of freedom available for such tests.

Ideally, since the basic variable concerns a total derived from household groups, the numbers in each group should be the same. It was not practicable to arrange matters so that this was the case (see Tables B and C). The variation in numbers will increase the variance of the basic variable, but the increase will only be small provided the coefficient of variation in numbers is small compared with that of expenditure. In so far as this additional source of variation is attributable to variation in response rate, it is acceptable and may be regarded as an integral part of the variation in the basic variable. On the other hand, variation due to differences in the numbers of households in various parts of the original sample cannot be so accepted, and requires correction.

In Table V the components of the variances in lines (*a*) to (*c*) have been derived from data based on the strata of Table A. The numbers in the latter table would have been the same along each row, and in the columns the numbers would have been in the ratios 5, 5, and 3 were it not for response variation. No adjustment has been made for this factor, but in calculating the between variances the basic data have been treated as though they had been derived from groups with frequencies proportional to 5, 5, and 3 for the three main strata respectively.

The same principle has been followed when computing the within variances. This consists of twelve components, one for each of the groups in Table A. The 40 groups in the larger urban strata contain about 20 households each, the 20 smaller urban groups about 40 each, and the rural groups about 60 each. Variances based on totals coming from these three sets of groups will be, approximately, in the ratio of 1, 2, and 3 respectively, and the appropriate adjustments were made in computing the within variance. On the other hand, no adjustment has been made for the slight variation between the numbers in the 20 groups of the smaller urban strata due to response (see Table B); the same applies to the variation between the numbers in the eight rural groups.

In the larger urban strata the variation in numbers between the 40 groups is greater than that arising from response variations. There was an additional source of variation due to the way the households in the four quarters of the stratum were divided into ten groups. This was done by allocating households into groups according to the final digit of the reference serial number, and as it turned out there were unequal numbers of households bearing these digits in the original sample. This additional variation could have been eliminated from the estimates of within variance by treating each group as though it had a frequency proportional to the numbers in the original sample. This would have produced estimates consistent with the other within variances, in that variation due to response would have been left in.

A trial was made using this correction for Table V and some of the categories in Table VI, but the difference made by the correction was so small that it was not worth proceeding with it. For consistency, Tables V, VI, and VII have been computed without this refinement. These tables involve variables y and z only, but the position was different with $\bar{x}$, the mean expenditure over the three weeks. The coefficient of variation for $\bar{x}$ is considerably smaller, in relation to that of the variation in numbers, than the corresponding coefficients for y or z, and the effect of the variation in numbers is thus much greater. The correction described in the previous paragraph has, therefore, been applied where $\mathrm{Var}(\bar{x})$ appears, that is in Table IV, and for consistency it has also been applied to the estimates of $\mathrm{Var}(y)$ and $\mathrm{Var}(z)$ in the same table.

REFERENCES

[1] Ministry of Labour (1957). *Report of an Enquiry into Household Expenditure in* 1953-54. HMSO, London.

[2] Kemsley, W. F. F. (1959). 'Designing a budget survey', *Applied Statistics*, **8**, 114.

[3] Kemsley, W. F. F. and Nicholson, J. L. (1960). 'Some experiments in methods of conducting family expenditure surveys', *J.R. Statist. Soc.*, A, **123**, 307.

[4] Turner, R. (1961). 'Inter-week variations in expenditure recorded during a two-week survey of family expenditure', *Applied Statistics*, **10**, 136.

INTER-WEEK VARIATIONS IN EXPENDITURE RECORDED DURING A TWO-WEEK SURVEY OF FAMILY EXPENDITURE

ROBERT TURNER

The author examines the extent to which, for one commodity (viz. tea), marked differences between the first week and second week records of individual households in a comprehensive family budget survey result in significant inter-week variation in the average per household of recorded expenditure on the commodity.

Introduction

In the Family Expenditure Survey (a continuing form of survey undertaken from the beginning of 1957 onwards by the Social Survey, Central Office of Information, on behalf of the Ministry of Labour), members of private households at a two-stage stratified sample of addresses in the United Kingdom are visited by interviewers and invited to keep fully itemised records of all their expenditure for a period of fourteen days. In each household in which every member who is working full-time or is 16 years of age or over agrees to keep records, a separate record book is provided for each spender for each of two consecutive seven-day periods. All spenders in a co-operating household keep records for the same period, but the survey is so conducted that, throughout the year, fresh households are visited each week and different households begin recording their expenditure on different days, normally days other than Saturday or Sunday. Information is sought during interviews about some types of expenditure, e.g. insurance, railway season tickets, electricity supply, over a period longer than two weeks. Information about income is also obtained. For convenience, the seven-day period covered by the first set of record books of a household is referred to as the *first week* of the survey of the household's expenditure and the following seven-day period as the *second week*, irrespective of the day on which the records begin.

Analysis of the results of the survey during 1957 and 1958 showed that, for various groups of items, the average per household of the expenditure recorded by households during the first week was significantly higher than the average of that recorded during the second week.[1]

Such inter-week variations were found in earlier surveys; for example, a series of small-scale experimental surveys in 1950 and 1951,[2] and the large-scale Enquiry into Household Expenditure in 1953-4.[3, 4] Variations of this kind have also been noted in surveys in other countries; for example, in Israel.[5]

These inter-week variations in expenditure recorded for some items indicated that the expenditure as recorded was abnormal in some way and that the results might be biased, but neither the kind of abnormality nor the direction of the bias, if any, in the mean weekly average expenditure was known. The actual expenditures might not have been recorded accurately; expenditure in the first week might have been overstated; expenditure in the second week might have been understated; some purchases might not have been recorded. Moreover the actual expenditures of the co-operating households during the survey might not have been representative of the expenditures of private households generally, because keeping a record book might have influenced their spending behaviour. Analysis of recorded expenditure unfortunately cannot by itself show how it differs from either actual or normal expenditure.

The fact that the results for different periods had this feature of inter-week variation in common indicated that it was not associated with specific periods of time or with households in specific districts surveyed in specific periods.

Within any particular field of expenditure, the households which recorded the same expenditure (possibly nil) in each of the two weeks would not give rise to any overall inter-week variation in expenditure in that field. The variation must arise from the differences between, on the one hand, the records of households recording higher expenditure on the item in the first week than in the second week, and, on the other hand, those of households recording lower expenditure (possibly nil) in the first week. If it could be shown that these groups of households had special characteristics, it might be possible to improve the survey techniques to reduce the inter-week variations in the survey results.

A Study of Expenditure on Tea

Scope. The present paper arises from a study in which special analyses were made, quarter by quarter in 1959, of the expenditure recorded on tea, and, to some extent, on other items on which significant inter-week variations in recorded expenditure had been found in the survey in both 1957 and 1958. 'Student's *t*', defined as the mean difference between the first and second week expenditures of all households co-operating in the survey divided by the standard error of the difference, had a highly significant value of nearly − 5 in both 1957 and 1958 for the item 'tea'.

This commodity was considered to be a suitable one for the purpose of this study as in nearly all households it is purchased by the housewife, normally in multiples of ¼ lb. units, fairly regularly but often not more than once a week. A marked difference in expenditures recorded by a household during the two survey weeks probably indicates a marked difference in the quantities purchased. Even if consumption within a household is fairly constant from week to week, however,

there may be variations in the quantities of this non-perishable commodity purchased by the household in different weeks, and no purchases may be made in some weeks, as larder stocks reach a certain level or for other reasons. Inter-week variations in expenditure by individual households are therefore to be expected. The two survey weeks in which records are kept are, however, independent of the normal tea-purchasing pattern of the household, and so any overall variation among a representative sample of households ought not to be significant.

The Classification of Households. In this survey,[1] as in the earlier Enquiry into Household Expenditure,[3] households in which at least three-quarters of the *total* income of the household is derived from National Insurance retirement or similar pensions and/or National Assistance paid in supplement or instead of such pensions, are regarded as a *special* group of households. Other households are classified according to the income of the head of the household and other characteristics.

For this study, households co-operating in the survey in 1959 were also classified according to whether, for this commodity, they recorded (*a*) no expenditure whatsoever in either week, (*b*) some expenditure in only the first of the two weeks, (*c*) some expenditure in only the second of the two weeks, or (*d*) some expenditure in both the first week and the second week. Households recording such expenditure in both weeks were further sub-divided according to whether or not there was a marked difference between the amounts recorded in the first and second weeks. In one quarter, similar classifications were made based on expenditure recorded on each of five other groups of items, viz. bread, potatoes, eggs, soap, and coal.

Average expenditure on tea varies considerably between households in different income groups, and so, to eliminate some of the effects of large variation in income, the study was concentrated on households in which the weekly income of the head of the household did not exceed £20 and excluded the *special* group of households. This main group of 2472 households included 80 per cent of the total number of households co-operating in the survey in 1959.

Analyses of Recorded Expenditure on Tea. The analyses of the survey data given in Tables I, II and III relate to this main group of households. An analysis of the recorded expenditure on tea of the excluded groups of households given in Table IV (at the end of this article) shows that, even though the average levels of their expenditure are very different, the main findings of the study are also applicable to these excluded groups of households.

The total expenditure on tea recorded by the 2472 households in the first week (9165 shillings) was nearly 7 per cent higher than the total in the second week (8576 shillings). Table I shows that, on average, 5 per cent of the co-operating households did not record any expenditure on tea in the two survey weeks, 10 per cent recorded ex-

TABLE I

Family Expenditure Survey, 1959—*analysis of expenditure on tea recorded by co-operating households in which the weekly income of the head of the household did not exceed £20*

Group of Households	Number of Households in the Group		Recorded Expenditure on Tea: Average per Household in the Group (Pence)		
	Number	Percentage of Total	First Week	Second Week	Mean
Households recording expenditure on tea in:					
(*a*) **Neither** week	120	5	—	—	—
(*b*) **First** week only	250	10	46·5	—	22·6
(*c*) **Second** week only	172	7	—	43·2	
(*d*) **Both** weeks	1930	78	51·0	49·5	50·2
(i) First week expenditure higher	522	21	67·3	39·3	53·3
(ii) Second week expenditure higher	433	18	42·5	69·8	56·1
(iii) No marked inter-week variation in expenditure	975	39	46·0	45·9	45·9
All households with first week expenditure higher (*b*) and (*d*)(i)	772	31	60·6	26·6	43·6
All households with second week expenditure higher (*c*) and (*d*)(ii)	605	25	30·4	62·2	46·3
All co-operating households	2472	100	44·5	41·6	43·1

penditure on tea only in the first week, 7 per cent recorded such expenditure only in the second week, and about 78 per cent recorded such expenditure in both weeks. There were marked differences between the first and second week records in half of the households recording expenditure on tea in both weeks.

Households Recording Expenditure on Tea in only ONE of the Two Weeks. The households recording purchases of tea only in the first week (250 or 10 per cent of the total) outnumber those recording purchases only in the second week (172 or 7 per cent). In these two groups, the mean weekly average expenditure per household (about 22½ pence) was a little under half that of households recording some expenditure in both weeks (about 50 pence). The difference (1·7 pence) between the mean weekly average expenditures in these two groups is not significant. (For all three thousand households co-operating in the survey, the standard error of the mean weekly average recorded expenditure on tea is about 0·5 pence.) The difference of 78 between the numbers of households in the two groups is, however, significant, having a standard error of approximately 20·5. This difference in number accounts for the total expenditure on tea recorded by the *first-week only* group (970 shillings) being substantially higher than that recorded by the *second week only* group (620 shillings). The records of these relatively small groups of households (17 per cent of the total) thus made a

proportionately large contribution (60 per cent of the total) to the overall inter-week variation in recorded expenditure on tea. These households do not appear to have any special general characteristics, e.g. size or composition of household, occupational status of the head of household, type of dwelling, type of area or geographical region, or to have been surveyed in particular weeks of the calendar quarters.

Households Recording Expenditure on Tea in Both Weeks. Among households recording such expenditure in both weeks, the total expenditure (and therefore the average expenditure per household) on tea was higher in the first week (8195 shillings) than in the second week (7956 shillings). These households, however, contributed less proportionately (239 shillings, or 40 per cent) to the overall inter-week variation (589 shillings) than did the households recording expenditure on tea in only one week (350 shillings). This relatively large group of households was sub-divided according to whether or not there was a marked difference between the expenditure on tea recorded in the first and second weeks. Slight differences (e.g. the difference between 3s. 6d. and 3s. 2d.) were assumed to reflect price or quality variations and not differences in quantities or purchasing habit. The sub-group of households recording higher expenditure in the first week (522 or over 21 per cent of all co-operating households) outnumbers the sub-group recording higher expenditure in the second week (433 or under 18 per cent). The excess of 89 is significant, having a standard error of approximately 30·9. This significant difference between the numbers of households in these two sub-groups accounts for the difference between the total first and second week expenditures being greater in the first sub-group than in the second sub-group. The mean weekly average expenditure per household was significantly higher in the second sub-group. In each of these sub-groups, the mean weekly average expenditure per household (about 53 and 56 pence respectively) was substantially higher than that of households recording approximately the same expenditure on tea in each of the weeks (about 46 pence).

The Level of Expenditure. An analysis of the expenditure on tea according to the level of expenditure of the household on tea is given in Table II. The households are grouped according to four bands of expenditure as follows:

(i) under four shillings in each week;
(ii) under six shillings in each week but four shillings or over in one week;
(iii) under ten shillings in each week but six shillings or over in one week;
(iv) ten shillings or over in one or both weeks.

Within each band, households recording expenditure on tea only in the first week outnumber those recording such expenditure only in the second week, have a higher total expenditure (except for the small

TABLE II

Family Expenditure Survey, 1959—*analysis of expenditure on tea according to the level of such expenditure recorded by co-operating households in which the weekly income of the head of the household did not exceed £20*

Groups of Households		Number of Households in the Group	Recorded Expenditure on Tea: Average per Household in the Group (Pence)		
Level of Recorded Expenditure on Tea	Households Recording Expenditure on Tea in:		First Week	Second Week	Mean
Nil in each week	(*a*) Neither week	120	—	—	—
Under 4 shillings in each week	(*b*) First week only	167	31·4	—	15·2
	(*c*) Second week only	133	—	29·6	
	(*d*) Both weeks:	1211			
	(i) First week expenditure higher	153	39·3	21·0	30·1
	(ii) Second week expenditure higher	121	21·0	39·0	30·0
	(iii) No marked variation in expenditure	637	. . .	. . .	32·9
Under 6 shillings in each week but 4 shillings or over in one week	(*b*) First week only	43	57·6	—	28·6
	(*c*) Second week only	19	—	56·4	
	(*d*) Both weeks:	585			
	(i) First week expenditure higher	169	58·9	36·1	47·5
	(ii) Second week expenditure higher	147	36·2	59·8	48·0
	(iii) No marked variation in expenditure	207	. . .	. . .	58·1
Under 10 shillings in each week but 6 shillings or over in one week	(*b*) First week only	32	84·7	—	42·6
	(*c*) Second week only	13	—	86·2	
	(*d*) Both weeks	467			
	(i) First week expenditure higher	175	87·7	53·7	70·7
	(ii) Second week expenditure higher	131	58·7	88·3	73·5
	(iii) No marked variation in expenditure	116	. . .	. . .	82·8
10 shillings or over in one or both weeks	(*b*) First week only	8	155·0	—	84·8
	(*c*) Second week only	7	—	186·4	
	(*d*) Both weeks:	89			
	(i) First week expenditure higher	25	152·7	73·3	113·0
	(ii) Second week expenditure higher	34	82·2	151·1	116·6
	(iii) No marked variation in expenditure	15	. . .	. . .	146·1

numbers of households in the highest expenditure band) than the latter and, in the lowest two bands, have a higher average expenditure per household. Their contributions to the overall inter-week variation were

(i) 106 shillings (ii) 117 shillings (iii) 132 shillings and (iv) minus 5 shillings.

Among households recording expenditure on tea in both weeks, within each band, the sub-group recording higher expenditure on tea in the first week is also larger in number than the sub-group recording higher expenditure in the second week (except in the highest band), and on average the inter-week patterns of expenditure in these sub-groups are exact reflections of one another, particularly in the large groups of households in the lowest two bands. For these two sub-groups combined, total expenditure recorded in the first week exceeds that recorded in the second week (except for the small numbers of households in the highest band). Their contributions to the overall inter-week variation were

(i) 52 shillings (ii) 35 shillings (iii) 174 shillings and (iv) minus 30 shillings.

Analysis of Expenditure on Certain Other Items

As tea-purchasing habits may be influenced by special factors peculiar to that item, recorded expenditure on various other items (viz. bread, potatoes, eggs, soap etc., and coal) was studied in a similar way in the third quarter of the year. These five other items include some perishable commodities which are purchased more frequently than tea and some non-perishable ones which are purchased less frequently.

The analysis of expenditure on these other items, given in Table III, has several features in common with that of expenditure on tea. The households recording expenditure on an item only in the first week outnumber those recording expenditure on the item only in the second week and, except for the very small numbers of households recording expenditure on bread in only one survey week, have a higher total expenditure. The mean weekly average expenditure per household of those households recording expenditure in one week only is less than half that of households recording expenditure on the item in both weeks, except for coal which, partly for seasonal reasons, was purchased by a much smaller number of households than the other items.

Conclusions

These analyses of expenditure show that, for various groupings of households, inter-week variations in recorded expenditure on tea are obtained. As these variations do not seem to arise from groups of households with general characteristics which can be specified, they may be an unavoidable general feature of expenditure survey data

TABLE III

Family Expenditure Survey, Third Quarter, 1959—analyses of expenditure on selected groups of items recorded by co-operating households in which the weekly income of the head of the household did not exceed £20

Item	Group of Households: Those Recording Expenditure on the Item in:	Number of Households in the Group	Recorded Expenditure on the Item: Average per Household in the Group (Pence)		
			First Week	Second Week	Mean
Bread	(*a*) **Neither** week	13	—	—	—
	(*b*) **First** week only	9	36·0	—	30·1
	(*c*) **Second** week only	9	—	84·6	
	(*d*) **Both** weeks	517	74·7	71·9	73·3
	All households	548	71·0	69·2	70·1
Potatoes	(*a*) **Neither** week	75	—	—	—
	(*b*) **First** week only	48	24·2	—	12·3
	(*c*) **Second** week only	28	—	25·2	
	(*d*) **Both** weeks	397	48·5	47·3	47·9
	All households	548	37·3	35·5	36·4
Eggs	(*a*) **Neither** week	55	—	—	—
	(*b*) **First** week only	70	51·7	—	23·4
	(*c*) **Second** week only	33	—	36·2	
	(*d*) **Both** weeks	390	57·7	56·2	57·0
	All households	548	47·6	42·2	44·9
Soap, detergents etc.	(*a*) **Neither** week	51	—	—	—
	(*b*) **First** week only	88	33·9	—	16·2
	(*c*) **Second** week only	59	—	30·2	
	(*d*) **Both** weeks	350	38·5	39·3	38·9
	All households	548	30·0	28·4	29·2
Coal	(*a*) **Neither** week	369	—	—	—
	(*b*) **First** week only	77	350·5	—	163·8
	(*c*) **Second** week only	38	—	280·9	
	(*d*) **Both** weeks	64	173·5	144·7	159·1
	All households	548	69·5	36·4	52·9
Tea	(*a*) **Neither** week	32	—	—	—
	(*b*) **First** week only	63	42·8	—	22·2
	(*c*) **Second** week only	35	—	47·1	
	(*d*) **Both** weeks	418	51·2	50·1	50·7
	All households	548	44·0	41·3	42·7

(Note: The Survey was suspended for part of this quarter immediately before the General Election and so the number of households covered is smaller than usual.)

obtained from spenders by record book methods. There are indications that this is so for some other items.

The relatively small group of households recording no expenditure on tea in one of the two survey weeks make a proportionately large contribution to the overall inter-week variation in recorded expenditure on this item. For every two households which record expenditure only in the second week there are about three which record such expenditure only in the first week. Some of these households may have omitted to record their purchases in one of the survey weeks. The mean weekly

average expenditure per household recorded by the *one-week only* households is consistently under half that recorded by other households. The difference between the numbers of *first-week only* and *second-week only* households is a much more imporant source of the inter-week variation than the difference between the average expenditures per household within these two groups. The mean weekly average expenditure on tea is higher, but not significantly, in the *first-week only* group than in the *second-week only* group.

Among the relatively large number of households recording expenditure on tea in both weeks, those recording markedly higher expenditure in the first week significantly outnumber those recording markedly higher expenditure in the second week, but these larger groups of households make a proportionately smaller contribution to the overall inter-week variation in recorded expenditure. Their mean weekly average expenditure per household is higher than that of those recording about the same expenditure in each of the two weeks. There is no marked inter-week variation in the expenditure on tea recorded by a substantial number of households (44 per cent).

Households with different levels of expenditure on tea contribute towards the overall inter-week variation, but those making larger purchases record more expenditure on tea in the second week.

The inter-week variations in the expenditure recorded may result from variations in actual expenditure: for example, if, whilst under survey, some households actually over-purchase in the first week and then either make a normal purchase or, possibly to offset their over-purchase, under-purchase or make no purchase in the second week. The extent to which purchasing habits can be adjusted in this way will vary between items, e.g. tea and bread. If such changes in habits occur and the over-purchasing in the first week is not balanced by the under-purchasing in the second week, the average over the two weeks of the expenditure recorded by these households will differ from their *normal* weekly expenditure. If the over-purchasing is balanced by the under-purchasing, these inter-week variations will not give rise to bias in the mean weekly average expenditure of these households.

If overall there are significant inter-week variations in the actual expenditure on tea by households co-operating in the survey, then the behaviour of these households is not representative, in this respect, of the behaviour of the class of households they are considered to represent. The fact that they are recording their expenditure for purposes of the survey may disturb their expenditure habits. By the methods being used in this survey, it may thus be impossible to obtain expenditure records which are free from this form of inter-week variation. Unless reliable comparable data can be obtained independently by different methods, it may be impossible to establish the reliability of the survey data and to determine whether the records of one week are more reliable than those of the other week.

The overall inter-week variations may arise, wholly or partly, from

inaccurate recording of actual expenditure by co-operating households. The recorded expenditure may be significantly higher in the first week than in the second week because, for some households:

(*a*) the first-week records overstate expenditure, possibly by including purchases actually made in the immediately preceding period; and/or

(*b*) the second-week records understate expenditure, possibly by omitting some purchases actually made.

The tendency to omit to record purchases may be stronger in the second week than in the first week. Inaccuracies in recording may not be intentional: for example, understatement of expenditure in the second week could be the effect either of fatigue in maintaining detailed and accurate records of expenditure—a task to which members of many private households are unaccustomed—or of a lowering of conscientiousness with which records are kept during the second week. A considerable and sustained effort may be required to keep completely accurate itemised records of expenditure over a period of fourteen days. If inaccuracies in recording occur, the reliability of the survey data, as a basis for estimates of the expenditure of private households, is dependent on whether or not the net overstatement of first-week expenditure by co-operating households differs from the net understatement of their second-week expenditure. This unfortunately cannot be determined from an examination of the survey data alone.

The reasons for these inter-week variations in expenditure recorded in family budget surveys are probably complex and have still to be explained. It is not easy to formulate alternative methods of collecting reliable information about family expenditure on such a comprehensive basis as in this survey, but it would be helpful if comparisons could be made with some data obtained independently, for example from retailers, for even a limited range of items.

Acknowledgements

This study was based on data extracted from Ministry of Labour records. I am grateful for permission to use the data and to publish this paper. I also wish to thank Mr W. F. F. Kemsley, Mr J. L. Nicholson, and the referee for comments they made on a preliminary version of the paper.

REFERENCES

[1] Ministry Of Labour (1961). *Family Expenditure Survey. Report for* 1957-59. HMSO, London.

[2] Kemsley, W. F. F. and Nicholson, J. L. (1960). 'Some experiments in methods of conducting family expenditure surveys', *J.R. Statist. Soc.*, A, **123**, 307.

[3] Ministry Of Labour (1957). *Report of an Enquiry into Household Expenditure in* 1953-54. HMSO, London.

[4] Kemsley, W. F. F. (1961). 'Household Expenditure Enquiry of the Ministry of Labour. Variability in the 1953-54 Enquiry', *Applied Statistics*, **10**, 117.

[5] Prais, S. J. (1958). 'Some problems in the measurement of price changes with special reference to the cost of living', *J.R. Statist. Soc.*, A, **121**, 312.

TABLE IV

Family Expenditure Survey, 1959—*analysis of expenditure on tea recorded by other groups of households and all households co-operating in the survey*

Group of Households		Number of Households in the Group	Recorded Expenditure on Tea: Average per Household in the Group (Pence)		
Class	Households Recording Expenditure on Tea in:		First Week	Second Week	Mean
Special (Pensioner) group	(*a*) **Neither** week	11	—	—	—
	(*b*) **First** week only	39	30·6	—	15·0
	(*c*) **Second** week only	17	—	29·3	
	(*d*) **Both** weeks	183	29·6	29·0	29·3
	All households	250	26·4	23·2	24·8
Weekly income of the head of the household over £20	(*a*) **Neither** week	42	—	—	—
	(*b*) **First** week only	53	59·6	—	30·6
	(*c*) **Second** week only	45	—	63·0	
	(*d*) **Both** weeks	230	52·7	48·8	50·7
	All households	370	41·3	38·0	39·6
All co-operating households	(*a*) **Neither** week	173	—	—	—
	(*b*) **First** week only	342	46·7	—	23·2
	(*c*) **Second** week only	234	—	46·0	
	(*d*) **Both** weeks	2343	49·4	47·8	48·6
	All households	3092	42·7	39·7	41·2

THE VARIABLE LOT-SIZE ACCEPTANCE SAMPLING PLAN FOR CONTINUOUS PRODUCTION

D. R. READ and D. W. BEATTIE
Cadbury Bros. Ltd., Bournville, Birmingham

In this article previously published acceptance sampling plans for continuous production are summarised and certain shortcomings pointed out. A new procedure is put forward in the Variable Lot-Size Acceptance Sampling Plan, which combines the useful features of lot-by-lot plans with the flexibility of continuous plans. The advantages and general features of this plan are described, and a table is given which enables AOQL values to be calculated.

Introduction

This paper is concerned with sampling inspection of units manufactured in quantity by an essentially continuous process, with sampling by attributes and a non-destructive test measurement.

The classical procedure, comprising rules for acceptance on a lot-by-lot basis, is not immediately applicable to continuous production. (The reader may note that the terms 'lot' and 'batch' are equivalent in the literature of acceptance sampling.) It is indeed possible to form 'lots' in some arbitrary manner, making them correspond to hours or shifts of production, for instance, but this artifice seldom has much to recommend it.

Several acceptance procedures designed specifically to deal with continuous production have been published during the last twenty years: but unfortunately they do not resolve certain practical difficulties, which will be described later. The object of this paper is to describe a new type of plan for continuous production, overcoming these difficulties.

A Summary of Previous Plans for Continuous Production

A brief review of previous plans will be given now, as a basis for later discussion.

The first plan, published by Dodge[1] in 1943 and later designated CSP-1, was as follows:

(*a*) At the outset, inspect 100% of the units consecutively as produced and continue such inspection until i units in succession are found clear of defects.

(*b*) When i units in succession are found clear of defects, discontinue 100% inspection, and inspect only a fraction f of the units, selecting individual sample units one at a time from the flow of product, in such a manner as to ensure an unbiased sample.

(*c*) If a sample unit is found defective, revert immediately to 100%

inspection of succeeding units and continue until again i units in succession are found clear of defects, as in paragraph (*a*).

(*d*) Correct or replace with good units all defective units found.

A series of curves was given to determine values of f and i for a given value of the Average Outgoing Quality Limit (AOQL).

Dodge and Torrey[2] subsequently introduced two modifications, designated CSP-2 and CSP-3: in principle these plans are the same as CSP-1; but instead of requiring reversion to 100% inspection whenever a defective is found during sampling inspection, CSP-2 and CSP-3 permit one or more defectives during sampling inspection, and call for reversion to 100% inspection only when the *spacing* between one defective and the next during sampling inspection is less than a prescribed minimum. CSP-3 includes a certain specific refinement to protect against surges of highly defective product.

Lieberman and Solomon[3] extended the Dodge plan CSP-1 to make for smoother transition between sampling and 100% inspection, the sampling fraction being changed from f^0 (100%) to $f^1, f^2, \ldots\ldots$, or back again, with some minimum f^k, according to the incidence of defectives or defective-free runs of length i units.

The referee has kindly pointed out that the U.S. Department of Defense Handbook H-106 describes a procedure that combines this multi-level approach with the special procedures of CSP-3.

A rather different type of plan was proposed by Wald and Wolfowitz,[4] thus:

(*a*) Choose a number N representing the number of units in a convenient segment of production.

(*b*) Start by inspecting a fraction f of the units, i.e. draw one item at random from each consecutive group of $1/f$ units.

(*c*) Submit the remainder of the segment to 100% inspection whenever, prior to the completion of the inspection of fN sample units, the number of defective sample units equals or exceeds

$$d = \frac{LN}{\frac{1}{f} - 1}$$

where L is a specified AOQL.

(*d*) Start afresh with each new segment of N units.

Girshick[5] put forward an alternative to the Wald-Wolfowitz scheme as follows:

(*a*) Inspection ends when a total of d defectives have been found.

(*b*) Start by inspecting a fraction f of the units, i.e., draw one unit at random from each consecutive group of $1/f$ units.

(*c*) When the dth defective has been found,

(i) the material passed is deemed satisfactory, and a new cycle of sampling is begun, if the number of groups examined (*g*)

is equal to or greater than some predetermined number G.

(ii) if $g<G$, then the following $G-g$ groups $[=(G-g)/f$ units] are inspected 100%, after which a new cycle of sampling inspection is begun.

If all defectives are replaced by good units, the above procedure gives an AOQL equal to $(1-f)d/G$; and if defective units are not replaced, the AOQL is d/G.

It is not the aim of the present paper to make a comparative assessment of these acceptance procedures; in any event a useful account has been given previously, in the readily available textbook by Duncan.[6]

Disadvantages of Previous Plans

These plans all suffer from the disadvantage of requiring an inspection rate that is liable to vary with time, even when the production rate is constant. With Dodge's CSP-1, for instance, if sampling inspection with $f=1/20$ requires one inspector, then 100% inspection will require about 20 inspectors; but extra inspectors—or extra personnel of *any* sort—cannot be procured at very short notice without involving the firm in considerable expense.

One possible way round this difficulty is to insert a buffer loop in the production line: this will enable the single inspector to 'fall back' in periods of 100% inspection and to catch up again in periods of sampling. Unfortunately this is seldom feasible on a real production lay-out, because space is usually restricted and the loop will have to be rather large to fulfil its purpose as a buffer; even then there will be an unduly delicate balance between low and high inspection rates, depending on the parameters of the acceptance plan and the pattern of incoming fraction defective.

At this juncture it seems desirable to emphasise that the variation is not in the *sampling* inspection rate (except with multi-level plans) but in the *total* inspection rate, i.e., the sampling inspection plus the occasional 100% inspection of rejected lots. Now the accepted practice regarding 100% inspection of rejected lots is that it should be performed, not by the Inspection Department, but by the Production Department, in order to make for more effective control. This is clearly still the right thing to do with continuous inspection; but the fact that the work load on the Inspection Department, by itself, is substantially constant does not in any way diminish the burden or cost of the variable work load on the Production Department. Albrecht *et al.*,[7] who have given an interesting description of continuous sampling procedure based on Girshick's plan, say (p. 4): 'We have made adjustments in the plan so that we can have an inspection force of constant size at all times. Most continuous plans involve bringing in extra inspection personnel at times of rejection'. Their procedure must nevertheless have involved extra work, and perhaps extra personnel also, in the Production Department at times of rejection; moreover this

extra work would have to be done more or less immediately, in order to dilute the fraction defective down to the AOQL and at the same time avoid patchiness of outgoing quality.

In all these continuous sampling plans, the occasional requirement for more or less immediate 100% inspection stems essentially from the need to dilute a run of bad product with a measure of good because a decision to do 100% inspection is reached *only after the run of bad product has been allowed to go through*. What is really required is a plan that will permit acceptance to cease at a point *before* the run of bad product starts.

The Variable Lot-Size Acceptance Sampling Plan

A suggested solution for this problem is the Variable Lot-Size Acceptance Sampling Plan, which combines the useful features of lot-by-lot plans with the flexibility of continuous plans. The plan is defined by a maximum sample size (n), an acceptance number (c), and a sampling fraction (f); its application is as follows:

(i) Designate any convenient point (i.e., unit of product) as the start of the first lot.
(ii) Inspect a fraction f of the units produced.
(iii) Start a new lot as soon as n units have been inspected, or as soon as $(c+1)$ defectives have been found, whichever occurs first.
(iv) If the sample contains c or fewer defectives, accept the lot. If the sample contains $(c+1)$ defectives, reject the lot: i.e., put it aside for subsequent 100% inspection.
(v) Replace all defective units found with good units.

It should be noted that this is not a continuous sampling plan (such as Dodge's CSP-1) and is merely intended to deal with the problem of continuous *production* where there may be no obvious 'lot' demarcation.

An essential feature of the plan is that the 100% inspection of rejected lots does not have to be done immediately but can be left for some hours or days, till spare labour becomes available; the continuous inspection rate is therefore constant. As regards the variation in quality from lot to lot, it is assumed that any pattern of variation resulting from this delayed 100% inspection is not of any practical consequence.

It will be observed that this plan does not suffer the great disadvantage of fixed lot-size plans in continuous production, viz., that the acceptance rule can lead to the rejection of units that have yet to be made. With the variable lot-size plan, decision is based only on the properties of, and only affects the fate of, units already produced. At the same time the results of variable lot-size samples can be used for quicker feedback of information on the persistence of bad quality or the point of return from bad quality to good quality.

In contrast with the continuous plans (i.e., those of Dodge, Wald and Wolfowitz, and Girshick), the variable lot-size plan results in a sentence (acceptance or rejection) that is carried out on the *same*

segment of product which was used to provide the decision-making sample: in other words it is retroactive, and avoids the situation where rejection (i.e. 100% inspection) is carried out only on the segment *following* that in which an excessive number of defectives was found (the rest of this apparently poor segment being *accepted*). With the variable lot-size plan, 100% inspection will tend to be applied to the bad patches themselves rather than to the patches that happen to follow the bad ones, which will occasionally be good ones.

The retroaction is sometimes awkward, of course, in that one has to insert a tag, or have some other suitable arrangement, to show where the lot started so that in the event of rejection the whole lot can be identified, collected, and put on one side for subsequent 100% inspection. This is rather a matter for inventiveness by local staff: in some instances it will present no problem at all; in others it will need careful thought. Then it will be necessary also to keep a count of the number of units inspected, in order to determine the end of the lot. For approximate control it may be sufficient to work on a time basis, assuming the inspection rate to be constant; for more accurate control a Veeder counter may be advisable.

The lot size varies, from $(c+1)/f$ to n/f. In this respect it is similar to the Girshick plan, which effectively has a variable lot-size (though almost infinitely variable), and it stands in contrast to the Wald and Wolfowitz plan, which has an effective fixed lot-size—the segment of N units.

Choice of a Variable Lot-Size Plan: The OC Curve, and the AOQL

The choice of a variable lot-size plan may be made in much the same way as the choice of a fixed lot-size plan: for instance, two quality levels, such as the Acceptable Quality Level (AQL) and the Lot Tolerance Percent Defective (LTPD), and associated probabilities of acceptance of lots having these qualities, can be specified. The OC curve for a variable lot-size plan with parameters f, n, c is in fact identical with the OC curve for a scheme with fixed lot-size n/f, sample size n, and acceptance number c. The OC curve referred to here is the Type B curve (Dodge and Romig,[8] p. 57) which gives the probability of accepting a lot as a function of *product quality* (rather than *lot quality*).

A choice based on the AQL and LTPD, or on the LTPD and minimisation of the total inspection effort, which is the Dodge-Romig scheme,[8] is not necessarily best from an economic point of view: but it is not the purpose of this paper to consider other criteria. An effective treatment of the economic problem, for rectifying inspection of continuous output, has already been made by Anscombe.[9]

It should be noted that, whereas in fixed lot-size plans the probability of a single item being accepted is identical with the probability of the lot being accepted, this does not hold for the variable lot-size plan where, in general, rejected lots are smaller than accepted lots. A

different type of OC curve (which we shall call Type C) could be drawn, in which the probability of an *item* being accepted is plotted as a function of product quality. Because rejected lots are smaller than accepted lots the AOQL, which is possibly the most useful criterion of effectiveness when considering plans applied to continuous production, is somewhat greater than that for the corresponding fixed lot-size plan. Fresh calculations are needed therefore in order to complete the numerical documentation of variable lot-size plans. The method of calculation, derived in the Appendix, is based on the relationship

$$\text{AOQ} = (1-f)\phi_A,$$

where the parameter ϕ_A is independent of f, and

$$\text{AOQL} = (1-f)\phi_L,$$

where ϕ_L is the maximum value of ϕ_A for variation of incoming quality (p). Values of $100\phi_L$ for $n=50(50)200$ and $c=0(1)4$ are shown in Table I.

TABLE I

Values of $100\phi_L$ *for variable lot-size plans*

Acceptance Number (c)	Sample Size (n)			
	50	100	150	200
0	1·30	0·65	0·43	0·32
1	2·24	1·12	0·75	0·56
2	3·28	1·64	1·09	0·82
3	4·42	2·21	1·47	1·10
4	5·61	2·80	1·87	1·40

Given n and c, and the maximum lot size (N) or the sampling fraction ($f=n/N$), the AOQL for any plan can easily be obtained. Thus, for example,

$$N = 600;\ n = 100,\ c = 0$$
$$100\phi_L = 0{\cdot}65$$
$$f = n/N = 1/6$$

Therefore $\text{AOQL} = (1-\frac{1}{6})(0{\cdot}65) = 0{\cdot}54\%$

Values of $100\phi_L$ in the corresponding fixed lot-size plans are given for comparison in Table II.

TABLE II

Values of $100\phi_L$ *for fixed lot-size plans*

Acceptance Number (c)	Sample Size (n)			
	50	100	150	200
0	0·74	0·37	0·25	0·18
1	1·68	0·84	0·56	0·42
2	2·74	1·37	0·91	0·69
3	3·88	1·94	1·29	0·97
4	5·09	2·54	1·70	1·27

The values in Tables I and II are based on the Poisson approximation to the binomial distribution.

The Total Amount of Inspection

It can be shown (see Appendix) that the total amount of inspection, i.e. the amount devoted to sampling plus the amount involved in 100% inspection of rejected lots, is less for the variable lot-size plan than for the equivalent fixed lot-size plan having the same values of n and c. The reason for this is best understood by considering a lot of size N divided into (say) ten sections, where the first section is of very poor quality and the remaining sections are all of good quality. In applying a fixed lot-size plan the whole lot may be condemned, and subjected to 100% inspection, as a result of the poor quality in the first section. With the variable lot-size plan, on the other hand, rejection and subsequent 100% inspection will be confined to the first section.

In more general terms we can say that, with the variable lot-size plan, poorer quality automatically leads to smaller lots, and therefore to less 100% inspection when runs of poor quality tend to be interspersed between runs of good quality, which is the very situation to which sampling inspection is normally applied.

Acknowledgements

We wish to acknowledge the computing assistance given by Miss M. C. Mason and the helpful comments made by the referee.

APPENDIX

Derivation of the OC Curve

The probability of acceptance of a lot is the probability that n items, drawn from a population where each item has an equal and independent chance of being defective, will show c or fewer defectives. This is precisely the probability given by a Type B OC curve.[8]

For any incoming flow of material with fraction defective p, there is a probability P_i that the $(c+1)$th defective will be the ith piece in the sample $(0 \leqslant i \leqslant n)$ and thus cause the lot to be rejected. Thus, P_i is equal to the probability that the first $(i-1)$ items include exactly c defectives, multiplied by the probability that the ith item is defective. That is,

$$P_i = \frac{(i-1)!}{c!\,(i-1-c)!}\, p^c (1-p)^{i-1-c} \times p$$

$$= \binom{i-1}{c} p^{c+1}\,(1-p)^{i-(c+1)}$$

The probability of acceptance *of a lot* (which will occur after n observations) is given by

$$P_a = 1 - \sum_{i=c+1}^{n} P_i$$

$$= 1 - p^{c+1} \sum_{i=c+1}^{n} \binom{i-1}{c} (1-p)^{i-(c+1)} \quad \ldots . (1)$$

The probability of acceptance *of an item* is given by

$$P_A = \frac{\text{Expected number of items accepted}}{(\text{Expected number of items accepted}) + (\text{expected number of items in a rejected lot})}$$

$$= \frac{\mathcal{N} P_a}{\mathcal{N} P_a + \bar{\mathcal{N}}_r}$$

where $\bar{\mathcal{N}}_r = \sum_{c+1}^{n} P_i \mathcal{N}_i$

$\mathcal{N}$ is the size of the lot of which n items are a sample, $\mathcal{N}_i$ is the size of a lot rejected at the ith observation, and

$$\mathcal{N}_i = \mathcal{N}i/n = i/f$$

where f is the fraction sampled.

In order to simplify the calculation of P_A, the following relation is needed:

$$\frac{1}{(c+1)} \sum_{c+1}^{n} i P_i = \left(\frac{1-p}{p}\right) \left\{1 - \mathrm{P_a}(n, c+1)\right\} + \left\{1 - \mathrm{P_a}(n, c)\right\} \quad \ldots . (2)$$

where $\mathrm{P_a}\ (n, j)$ = probability of acceptance of a lot when a sample of size n is taken and the acceptance number is j.

The relation (2) follows simply on the substitution of $\mathrm{P_a}\ (n, c)$ and $\mathrm{P_a}\ (n, c+1)$ from equation (1).

Therefore

$$P_A = \frac{n\ \mathrm{P_a}\ (n, c)}{(c+1) + (n-c-1)\ \mathrm{P_a}\ (n, c) + (c+1)\ (1-p)\ \{1 - \mathrm{P_a}(n, c+1)\}/p}$$

The Average Outgoing Quality

For any given input quality p, the AOQ is given by

$$\mathrm{AOQ} = \frac{(\text{Number of defectives in an accepted lot})(P_a)}{(\text{Number of items in an accepted lot})(P_a) + (\text{Expected number of items in a rejected lot})}$$

$$= \frac{(\mathcal{N}-n)\ p\ P_a}{\mathcal{N} P_a + \bar{\mathcal{N}}_r},$$

$$= \frac{(\mathcal{N}-n)\ p\ P_a}{\mathcal{N} P_a + \sum_{c+1}^{n} P_i \mathcal{N}_i}$$

In deriving the AOQ it has been assumed that defectives found in the sample, as well as all defectives in rejected lots, have been rectified. If, however, we assume for the moment that defectives found in the sample are *not* rectified, we have an expression for the AOQ,

$$\phi_A = \frac{N p P_a}{NP_a + \sum_{i=c+1}^{n} P_i N_i} = \frac{p P_a}{P_a + \frac{1}{n}\sum_{c+1}^{n} i P_i} = p P_A$$

The importance of the parameter ϕ_A is that it is independent of the fraction sampled (and of the lot size) so that, given ϕ_A, the AOQ when defectives in the sample are rectified can readily be found for any specified value of f, thus:

$$\text{AOQ} = (1-f)\phi_A$$

Similarly $\text{AOQL} = (1-f)\phi_L$

where $\phi_L = \max(\phi_A)$.

For the fixed lot-size plan we have

$$\text{AOQ} = \frac{(N-n)p\,P_a}{N}$$

Therefore $\phi_A = p\,P_a$

and this is obviously less than ϕ_A for the variable lot-size plan because

$$P_a + \frac{1}{n}\sum_{c+1}^{n} i P_i = 1 - \sum_{c+1}^{n} P_i + \frac{1}{n}\sum_{c+1}^{n} i P_i < 1$$

Using the Poisson approximation,

where $$\mathrm{P_a}(n, c) = \sum_{m=0}^{c} \frac{e^{-np}(np)^m}{m!}$$

we can show that $$\phi_L = \frac{1}{n} y'$$

where y' is independent of n. Therefore for a given value of c, ϕ_L can be calculated for any value of n, within the limits for which the Poisson approximation is adequate.

The Total Amount of Inspection

The total amount of inspection is the sum of two parts: the amount of sampling and the amount of 100% inspection of rejected lots. Thus

$$\text{Total Inspection (per item produced)} = \frac{nP_a + \bar{N}_r}{NP_a + \bar{N}_r}$$

For the fixed lot-size plan with the same n and c, the total amount is $1 - P_a(1-f)$.

It can easily be shown that

$$\frac{nP_a+\bar{N}_r}{NP_a+\bar{N}_r} \leqslant 1-P_a(1-f)$$

the equality referring to the extreme case $P_a=1$. In other words, the total amount of inspection is less for the variable lot-size plan than for the equivalent fixed lot-size plan.

REFERENCES

[1] Dodge, H. F. (1943). 'A sampling inspection plan for continuous production', *Ann. Math. Statist.*, **14**, 264.

[2] Dodge, H. F. and Torrey, M. N. (1951). 'Additional continuous sampling inspection plans', *Ind. Qual. Control*, **7** (5), 7.

[3] Lieberman, G. J. and Solomon, H. (1955). 'Multi-level continuous sampling plans', *Ann. Math. Statist.*, **26**, 686.

[4] Wald, A. and Wolfowitz, J. (1945). 'Sampling inspection plans for continuous production which insure a prescribed limit on the outgoing quality', *Ann. Math. Statist.*, **16**, 30.

[5] Girshick, M. A. (1948). 'Sampling inspection plans for continuous production'. Unpublished paper delivered at the meeting of the Institute of Mathematical Statistics, 10 May 1948.

[6] Duncan, A. J. (1959). *Quality Control and Industrial Statistics*. Irwin, Illinois.

[7] Albrecht, L., Gulde, H., Maclean, A. and Thompson, P. (1955). 'Continuous sampling at Minneapolis-Honeywell', *Ind. Qual. Control*, **12** (3), 4.

[8] Dodge, H. F. and Romig, H. G. (1959). *Sampling Inspection Tables. Single and Double Sampling*, 2nd edition. Wiley, New York.

[9] Anscombe, F. J. (1958). 'Rectifying inspection of a continuous output', *J. Amer. Statist. Ass.*, **53**, 702.

EMPIRICAL EVIDENCE ON OPTIMUM MOTOR TAXES FOR THE UNITED KINGDOM

A. A. WALTERS
Birmingham University

Using the results of a Road Research Laboratory survey of traffic conditions in Slough High Street the author questions the level of motor taxation. The application of multiple regression techniques to these data, coupled with information derived elsewhere, suggests that the optimum level of taxation may be somewhat higher than the present level.

Introduction

In this paper an attempt is made to estimate the level of optimum taxes for motor vehicles in the United Kingdom. The theory of efficient motor taxes goes back at least 40 years to Pigou.[1] Pigou argued that road users take account only of their private cost in deciding which road to use. But since the motorist adds to the congestion and so to the costs of *all* road users, marginal *social* costs exceed marginal private costs. Pigou suggested that vehicles using congested roads should be taxed by the difference between social and private cost. Recently the theory has been refined and extended using modern programming techniques.[2] However, the application of the Pigou theory has hardly been attempted. A preliminary investigation of American data using the Pigou model suggested that the level of motor taxes should be greatly increased.[3] In addition it was argued that special imposts such as mileage taxes should be levied in very congested cities.

Now what of the United Kingdom? The main suggestion that emerges from the following analysis is that there is no case for reducing petrol taxes. There is a rather inconclusive case for *raising* petrol taxes, which might be reinforced if more basic data were accumulated. The analysis is also used to assess the marginal cost of parking a vehicle for an hour. The results suggest that the parking charges should be of the order of 1s. 6d. to 3s. an hour if they are to reflect the marginal social cost. An attempt is made to extend the analysis by allowing for the feedback effect of reduced traffic flow on pedestrians and so on the speed (costs) of traffic. Little progress, however, can be made along these lines until data become available for the analysis of a more extensive system of equations.

In this essay on applied welfare economics many of the t's must remain uncrossed and the i's undotted. In particular we shall not be concerned with the effects on the distribution of income. Secondly we shall suppose that all commodities and services closely competitive with road services are produced under conditions near enough to

perfect competition. Public transport is one obviously close substitute which can rarely be approximated by a perfect competition model. But it seems likely that prices of public transport services are as often above marginal cost as below. Thirdly we shall normally assume that there is a fixed ratio between petrol consumption and vehicle mileage, independent of relative prices. Other assumptions specifically concerned with the empirical material will be dealt with as they arise.

Theory

When the individual motorist decides to undertake a journey he assesses only the *private* marginal cost. But when he joins other vehicles on the road he adds to congestion. All other motorists suffer an increase in costs due to this additional congestion. Thus marginal social costs exceed marginal private costs unless the additional vehicle does not impede and is not itself slowed down by the existing traffic.

Assuming traffic consists of homogeneous units, we write aggregate private costs of vehicles (K) as a function of traffic (X):

$$K = \mathrm{K}(X)$$

and $\dfrac{K}{X} = C$ is the marginal private cost of each vehicle.

Then
$$\frac{\mathrm{d}K}{\mathrm{d}X} = \frac{\mathrm{d}(CX)}{\mathrm{d}X}$$

$$= C\left(1 + \frac{X}{C} \cdot \frac{\mathrm{d}C}{\mathrm{d}X}\right) \qquad \ldots\ldots (1)$$

or, in words:

$$\text{marginal social cost} = \text{marginal private cost} \times (1 + \text{elasticity of marginal private cost}).$$

Writing the inverse of the demand function (the 'marginal valuation curve') as $P = \mathrm{P}(X)$ where P is the marginal valuation of an additional journey when there are X vehicle journeys, we have the market equilibrium solution:

$$\mathrm{P}(X) = \mathrm{C}(X)$$

The ideal quantity of traffic is the solution of

$$\mathrm{P}(X) = \frac{\mathrm{d}K}{\mathrm{d}X}$$
$$= C(1 + \eta_X)$$

where η_X is the cost elasticity. We seek methods of estimating η_X and imposing a tax or toll of $C\eta_X$.

The normal shape of the $\mathrm{C}(X)$ curve is upward sloping from left to right, Fig. 1.

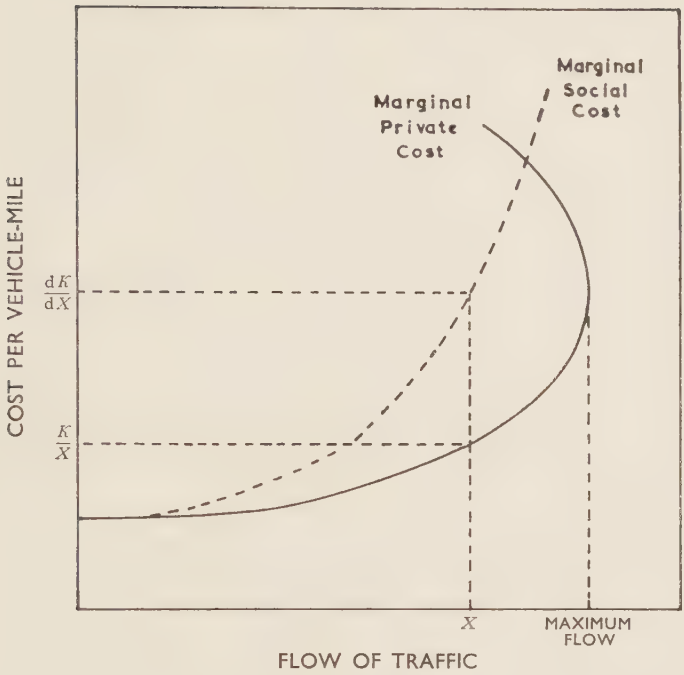


Fig. 1. Dependence of motoring costs upon traffic flow.

As traffic increases so marginal private costs increase. But in very congested traffic conditions, additional vehicles may increase the density to such an extent that the actual *flow* of vehicles (i.e. the number of vehicles a minute passing a point on the road) is diminished. This means that, with given road capacity, there is a backward bending 'supply curve' of road services, with a *maximum* flow as shown in Fig. 1. This is not a conventional kind of supply curve; for high demand, leading to higher density, reduces flow (for a fuller discussion see reference 3). Curves of this shape have been shown to be a good approximation to traffic flows on certain high-density roads in the USA. But for low-density highways a simple linear approximation has been shown to be useful. We now proceed to examine the empirical data for the United Kingdom.

Empirical Evidence

The main data used to estimate cost elasticities were collected by the Road Research Laboratory.[4] In Slough High Street during 1952, 429 observations were taken of the number of motor vehicles on the road, the number parked, and the number of pedestrians on the road. The investigators in fact collected and analysed much more detailed data—including the direction of traffic, the various classes of vehicle, the number of pedestrians on crossings, and the number off crossings, for the Laboratory was mainly interested in the effect of changes in

waiting restrictions. These data do not include figures of cost. But, as a first approximation at least, we assume that marginal private cost is inversely proportional to speed. If this assumption is near enough correct, the estimate of the elasticity of speed with respect to traffic will be a good estimate of the elasticity of marginal private cost.

The data were plotted to see the form of the variation of speed with the independent variables. There was no marked evidence of curvilinearity. In particular, it did not seem that there was any reason to suppose that the effective part of the curve was 'backward bending'. Thus the data were fitted to a simple regression model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \epsilon_t \quad \ldots . \quad (2)$$

where X_{1t} is the traffic flow (East and West) in numbers of vehicles (excluding pedal cycles) an hour in period t.
X_{2t} is the number of parked vehicles on both sides of the road.
X_{3t} is the number of pedestrians off crossings.
X_{4t} is the number of pedestrians on crossings.
Y_t is the speed of light cars and vans.
ϵ_t is a random disturbance with zero mean.

The estimates together with the confidence intervals are given in

TABLE I

Estimated regression coefficients of equation (2)

$\bar{R}^2 = 0{\cdot}711$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
Estimated coefficient	−0·006 924	−0·031 169	−0·000 060	−0·000 932
95 per cent confidence interval	−0·005 644	−0·024 818	+0·000 131	−0·000 771
	−0·008 204	−0·037 520	−0·000 251	−0·001 094

Table I. (The correlations between these estimated coefficients were generally low, except for the case of β_3 and $\hat{\beta}_4$, suggesting that the regression coefficients are fairly stable ones.)

The elasticities of speed with respect to the independent variables are not constant. For example, as traffic flow increases the absolute value of the elasticity increases. We must therefore specify the values of the independent variables before calculating elasticities. With $X_1 = \bar{X}_1$, $X_2 = \bar{X}_2$, $X_3 = \bar{X}_3$, we get an elasticity of speed with respect to traffic flow as follows (here, and subsequently, the subscript t has been omitted since no confusion is likely):

$$(\eta_{X_1})_{\bar{X}} = \frac{\bar{X}_1}{\bar{Y}} \cdot \frac{\partial Y}{\partial X_1} \quad \ldots . \quad (3)$$

where $\bar{Y}$ is the estimate of the average speed. This is the *partial* elasticity of speed with respect to traffic flow and is a useful concept, in practice, if it is possible to charge efficient prices to pedestrians and parked vehicles. If, however, administrative instruments for pricing pedestrians and parkers are lacking, the *total* elasticity is the more relevant

concept. The corresponding estimator of the partial elasticity from the regression is:

$$(\eta_{X_1})_{\bar{X}} = \frac{\bar{X}_1}{\bar{Y}} \cdot \hat{\beta}_1 \qquad \ldots . (3)$$

In Table II the elasticities are calculated for the average values of the independent variables ($\bar{X}_r$) and for $\bar{X}_r - \sigma_r$ and $\bar{X}_r + \sigma_r$ ($r = 1, 2, 3.$) where σ_r is the standard deviation of X_r.

TABLE II

Partial elasticities of speed (Y) derived from the estimated equation (2)

Value of the Independent Variable (X_r)	Partial Elasticity of Speed (Y) with Respect to:			
	Vehicle Flow (X_1)	Parked Vehicles (X_2)	Pedestrians off Crossings (X_3)	Pedestrians on Crossings (X_4)
$\bar{X}_r - \sigma_r$	−0·318	−0·050	−0·008	−0·092
$\bar{X}_r$	−0·419	−0·099	−0·014	−0·195
$\bar{X}_r + \sigma_r$	−0·563	−0·167	−0·023	−0·343

The most important elasticities are those with respect to traffic flow. At average traffic flow, a one per cent change in vehicles an hour gives rise to a 0·42 per cent decrease in speed, if the numbers of parked vehicles, and of pedestrians on and off crossings, remain constant. The values in the table are rather low compared with data for similar urban traffic in the United States (considered in more detail in reference 3).

In order to apply the theory outlined earlier we should also know the elasticity of the demand curve. Unfortunately, there is very little data available on demand conditions, and the analysis of a demand which fluctuates violently from one ten-minute period to another is exceedingly difficult. Consequently no attempt has been made in this paper to estimate elasticities of demand. This severely limits the statements we can make about optimum tax policy. But from estimates of cost elasticities and assuming that the law of demand applies we can certainly predict the *direction* in which taxes should move. In addition we can usually make an informed guess at the order of magnitude of the change.

Optimum Taxes

We can now calculate optimum taxes for any given level of traffic. There is however, a preliminary point to be cleared up. Instead of measuring cost ($C = K/X_1$) as the independent variable in our regression equation we have estimated speed (Y). We assume that speed is inversely proportional to marginal private cost:

$$Y = \frac{k}{K/X_1} \text{ (where } k \text{ is a constant)}$$

Substituting into equation (2), and ignoring the disturbance, we get:

$$k\left(\frac{X_1}{K}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \qquad \ldots . (4)$$

Differentiating partially with respect to X_1 we get:

$$k\left(\frac{1}{K}-\frac{X_1}{K^2}\frac{\partial K}{\partial X_1}\right)=\beta_1$$

$$\text{i.e. } \frac{\partial K}{\partial X_1}=\frac{K}{X_1}\left(1-\beta_1\frac{K}{k}\right)$$

$$=C\left(1-\beta_1\frac{X_1}{Y}\right) \qquad \ldots\ldots (5)$$

where $\beta_1 X_1/Y$ is the elasticity of speed with respect to vehicle flow. Equation (5) corresponds exactly to equation (1). If marginal private cost is inversely proportional to speed, the elasticity of speed with respect to traffic flow (with a change in sign) will serve to measure the cost elasticity.

We now require estimates of K/X_1 and k. Various accounts of motoring costs at various speeds are available. Probably one of the most careful assessments is by G. J. Roth.[5] We have adjusted Roth's estimate by (i) deducting the element of depreciation cost since this depends primarily on time and not on mileage, and (ii) adding the time cost of driver and passengers. This gave an average cost of 10d. a vehicle mile for a speed of 20 m.p.h.

Then: $$k = 10 \times 20 = 200$$

Inserting these values in the above formula we get the results of Table III.

The optimum tax of between $3\frac{1}{2}$ pence and 9 pence a mile must be compared with the existing tax of about 1·1 pence to 1·6 pence a mile. This suggests that there is a good case for increasing taxes in traffic conditions similar to Slough High Street.

TABLE III

The estimated optimum tax for a given level of traffic, derived from equation (5)

Level of Traffic (Vehicles per Hour)	Average Cost (Pence per Mile)	Optimum Tax (Pence per Mile)	Speed (m.p.h.
X_1	K/X_1	$\hat{\eta}_{X_1}\cdot\frac{K}{X_1}$	Y
$\bar{X}_1-\sigma_1$	10·5	3·4	19·04
$\bar{X}_1$	12·5	5·3	16·03
$\bar{X}_1+\sigma_1$	15·4	8·7	13·02

The results so far have been based on *partial* elasticities. They tell us what the optimum tax should be, if it were also possible to charge efficient prices or taxes for parking and pedestrians. In fact the figures in Table III refer to the *existing* amount of parking and pedestrians. It is of course quite conceivable, and I think also desirable, that efficient prices should be charged for road-side parking. We return

to this in the next section. But as for pedestrians, it seems administratively impossible to charge suitable prices for the privilege of walking across streets. We might then calculate what the efficient tax should be *taking into account* the reactions of pedestrians, but still assuming that the behaviour of parkers can be controlled by the price system. The usual technique for tackling this problem would be to develop a system of equations and estimate the structural parameters. Obviously such a task is beyond our data. However, we tried an alternative one-equation approach to see if it would give an estimate of the gross or total marginal cost when only X_2 is fixed and when we take into account variations of the number of pedestrians.

One approach is to differentiate equation (4) with respect to X_1 assuming that X_2 is fixed. Thus, if we write the total differential $\left(\frac{\mathrm{d}}{\mathrm{d}X_1}\right)_{X_2}$ as the differential coefficient when only X_2 is fixed we get:

$$k\left[\frac{1}{K}-\frac{X_1}{K^2}\left(\frac{\mathrm{d}K}{\mathrm{d}X_1}\right)_{X_2}\right]=\beta_1+\beta_3\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}+\beta_4\left(\frac{\mathrm{d}X_4}{\mathrm{d}X_1}\right)_{X_2} \quad \ldots\ldots (6)$$

$$\text{i.e. } \left(\frac{\mathrm{d}K}{\mathrm{d}X_1}\right)_{X_2}=\frac{K}{X_1}\left[1-\frac{K}{k}\left\{\beta_1+\beta_3\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}+\beta_4\left(\frac{\mathrm{d}X_4}{\mathrm{d}X_1}\right)_{X_2}\right\}\right]$$

$$=C\left[1-\frac{K}{k}\left\{\beta_1+\beta_3\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}+\beta_4\left(\frac{\mathrm{d}X_4}{\mathrm{d}X_1}\right)_{X_2}\right\}\right] \ldots\ldots (7)$$

The limitations of this approach must be clearly recognised. It treats the total cost as though it arose from one exogenous influence, i.e. traffic flow. The effect of this variable is transmitted either directly, or indirectly through its effect on pedestrians.

To calculate $\left(\frac{\mathrm{d}K}{\mathrm{d}X_1}\right)_{X_2}$ from equation (7) we require estimates of $\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}$ and $\left(\frac{\mathrm{d}X_4}{\mathrm{d}X_1}\right)_{X_2}$. These are the changes in the numbers of pedestrians due to a change in the volume of traffic when the number of parked cars is fixed. The natural way of estimating a linear approximation to $\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}$ is to form the regression equation of X_3 on X_1 when the effect of X_2 has been eliminated. Thus, in Yule's notation, we calculate the regression:

$$x_{3.2}=\beta_{31.2}x_{1.2}$$

or

$$x_3=\beta_{0.12}+\beta_{31.2}X_1+\beta_{32.1}X_2$$

The value $\hat{\beta}_{31.2}$ is then an estimate of $\left(\frac{\mathrm{d}X_3}{\mathrm{d}X_1}\right)_{X_2}$. Similarly the value $\hat{\beta}_{41.2}$ is an estimate of $\left(\frac{\mathrm{d}X_4}{\mathrm{d}X_1}\right)_{X_2}$. The data give the values shown in Table IV.

TABLE IV

Estimated partial regression coefficients of pedestrians on (X_3), and off (X_4), crossings with respect to traffic flow (X_1) and the number of parked cars (X_2)

Dependent Variable	Traffic Coefficients		Parking Coefficients	
	Estimated Coefficient	95% Confidence Interval	Estimated Coefficient	95% Confidence Interval
Pedestrians off crossings (X_3)	−2·44 ($\hat{\beta}_{31.2}$)	−1·36 −3·53	21·08 ($\hat{\beta}_{32.1}$)	26·62 15·54
Pedestrians on crossings (X_4)	0·54 ($\hat{\beta}_{41.2}$)	1·83 −0·74	26·51 ($\hat{\beta}_{42.1}$)	33·06 19·96

Inserting these values in equation (7) we get the marginal cost $\left(\frac{dK}{dX_1}\right)_{X_2} = 18$ pence for average values of the independent variables. This is about the same as the computed value of $\frac{\partial K}{\partial X_1}$, namely 17·8 pence.

One trouble with the total marginal cost approach is that it incorporates certain 'net' effects from the complete system underlying this single equation. The negative coefficient $\beta_{31.2}$, for example, indicates that there are more 'off-crossing' pedestrians the lower the volume of traffic, but it is unlikely that an increase in the price of motoring will give rise to lower traffic volume which in turn will induce more 'off-crossing' pedestrians which *then* will lower the speed of vehicles. But this is the implication of equation (7). It is much more plausible to suppose that the causal relation is the other way—pedestrians cross the street when they can find a gap in the traffic flow. Furthermore our model ignores the cost to pedestrians of waiting for a gap in traffic or walking to a crossing. All these considerations would suggest that the true marginal social cost is above the computed value $\left(\frac{dK}{dX_1}\right)_{X_2}$. But a more complicated analysis to discover these side effects must await the collection of detailed data.

A modest adjustment can be made to our existing model by *excluding* pedestrians 'off-crossings' from the simple regression model. We argue that the causal relationship is from traffic *to* 'off-crossing' pedestrians and that we can ignore the feedback effects. The results upon estimating this relation are shown in Table V.

TABLE V

Estimated regression coefficients derived from equation (2), *after excluding pedestrians off crossings (X_3)*

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_4$
Estimated coefficient	−0·00675	−0·03129	−0·00098
Elasticity at average values	−0·409	−0·099	−0·204

These values (Table V) are about the same as those derived from the regression with four independent variables (Table I). The omission

of pedestrians off crossings makes little difference to the estimates of elasticity.

Optimum Parking Charges

The earlier regression analysis (p. 160) can easily be made to yield estimates of the marginal cost of parking a car for an hour. As in the preceding section, we first calculate the effect on speed of a small increase in number of cars parked while the number of pedestrians and volume of traffic remain constant. Thus we differentiate equation (4) partially with respect to X_2:

$$k \cdot \frac{X_1}{K^2} \cdot \frac{\partial K}{\partial X_2} = \beta_2$$

$$\frac{\partial K}{\partial X_2} = \frac{K^2 \beta_2}{k X_1}$$

$$= \frac{\beta_2}{Y} \cdot K \qquad \ldots (8)$$

The marginal cost of a vehicle-hour of parking is composed of the product of two factors: (i) the decrease in speed for a small increase in vehicles parked (β_2), divided by average speed; and (ii) the total cost of the existing traffic flow.

Applying this formula to average conditions on Slough High Street, we get:

$$\left(\frac{\partial K}{\partial X_2}\right)_{\bar{X}_r} = \frac{k \bar{X}_1}{\bar{Y}^2} \cdot \hat{\beta}_2 \qquad r = 1, 3, 4 \qquad \ldots . (9)$$

where $\bar{Y}$ is the mean speed and $\hat{\beta}_2$ is the estimate of the regression coefficient β_2.

$$\left(\frac{\partial K}{\partial X_2}\right)_{\bar{X}_r} = \frac{200 \times 917{\cdot}1}{16{\cdot}03^2} \times 0{\cdot}36196$$

Marginal cost for a vehicle-hour of parking = 26 pence approximately.

This can be interpreted as an estimate of the optimum parking-meter charge in conditions of average traffic flow, etc. The actual charge should, of course, be lower than this when demand conditions are taken into account. We have no information on the elasticity of demand for parking.

If we suppose that the government can devise suitable approximations to optimum taxes for road traffic, but that, on the other hand, it is administratively impossible to charge pedestrians efficient prices, equation (8) becomes

$$\left(\frac{dK}{dX_2}\right)_{X_1} = -\frac{K^2}{kX_1}\left[\beta_2 + \beta_3\left(\frac{dX_3}{dX_2}\right)_{X_1} + \beta_4\left(\frac{dX_4}{dX_2}\right)_{X_1}\right] \qquad \ldots . (10)$$

Using $\hat{\beta}_{32.1}$ and $\hat{\beta}_{42.1}$ as estimates of $\left(\frac{dX_3}{dX_2}\right)_{X_1}$ and $\left(\frac{dX_4}{dX_2}\right)_{X_1}$, we get

$$\left(\frac{dK}{dX_2}\right)_{X_1} = 43 \text{ pence an hour compared with}$$

$$\frac{\partial K}{\partial X_2} = 26 \text{ pence.}$$

In this case consideration of the secondary effects on pedestrians has increased considerably the optimum parking charge. But probably little confidence can be placed in this value because it is unlikely that, apart from traffic flow, the number of parked vehicles is the only independent variable which exerts a substantial effect on the number of pedestrians. Indeed one may doubt whether the secondary effect represents a causal relationship at all. Until the causal relations can be sorted out and measured it is probably best to ignore the secondary effects in formulating policy.

Additional Evidence and the Applicability of the Results

So far our results refer to only urban traffic conditions. Broadly speaking there are three other classes of traffic: main trunk roads in rural areas, subsidiary rural roads, and subsidiary roads in urban districts. Only fragmentary evidence is available on the speed-volume relationship and very little on the cost-volume function.

For rural roads with light traffic in Buckinghamshire the Road Research Laboratory carried out a multiple regression analysis of speed on traffic flow, width of carriageway, curvature, steepness, number of access points per mile.[6] The regression coefficient on traffic flow was not significantly different from zero. This suggests that marginal social cost and marginal private cost are equal to a constant. Optimum taxes are zero.

For rural trunk roads the evidence of Coburn[7] is most useful. For two-lane stretches of A5 Coburn showed that the mean speed of private cars (Y) was related to total traffic flow (X) by the equation:

$$Y = 41{\cdot}9 - 0{\cdot}119X$$

and for the three-lane stretches the equation was:

$$Y = 47{\cdot}7 - 0{\cdot}0087X$$

At mean daytime traffic flow, the regression coefficient for the two-lane stretch implies an elasticity of speed with respect to traffic flow of about $-0{\cdot}13$. The cost elasticity is probably less than $-0{\cdot}10$ and this implies an optimum tax of about $\frac{1}{2}$d. a mile. This is about half the existing tax rate of about a penny a mile. In the extreme case of motorway M1, it is certain that the cost elasticity is zero and that there should be no tax.

What is the distribution of vehicle mileage over these various kinds of roads? There are, of course, no statistics of traffic classified according

to the elasticity of speed. But a recent study[8] has shown that 25-30 per cent of vehicle mileage is on the busiest 1 per cent of the road mileage, 60 per cent on the busiest 10 per cent, and 95 per cent on the busiest 50 per cent. This shows a marked concentration of traffic on urban and trunk roads. It seems likely that much traffic will experience the high cost elasticities and comparatively little the zero elasticities. But this is, of course, merely a conjecture; direct evidence on this point is still needed.

The elasticities calculated above (p. 161 and p. 166) refer to traffic conditions in 1952. In applying these results in the 1960's, it must be remembered that between 1951 and 1958 traffic has increased by an average of 7 per cent a year, and between 1958 and 1959 the increase was 12 per cent. We do not know in detail how this increase was distributed according to the cost elasticity, although there is evidence that in Central London the increase in night traffic was much greater than that in daytime traffic.[9] But it is reasonable to suppose that this large increase in traffic combined with the very small increase in urban road capacity has given rise to an *increase* in the cost elasticities. Optimum taxes in the 1960's are clearly higher than those for 1952.

Fuel Taxes and Tolls

The calculations we have made of optimum taxes are rough and ready guides to policy. In implementing that policy we have to use very blunt instruments indeed, namely motor taxation and tolls. The basic difficulty is that motor taxation can discriminate to only a very slight extent between congested and empty roads. Taxes on fuel per mile vary because fuel consumption increases as congestion increases (for speeds below about 35-40 m.p.h.), but this variation could only be used to reflect the values of the optimum taxes on congested roads by imposing a very large fuel tax indeed. This would have many undesirable side effects—under-utilisation of rural roads, and (socially) uneconomic engine designs in pursuit of low petrol consumption. The main objections to tolls are that they are apparently expensive to collect and cause congestion by their very collection. Various other measures include special licence stickers for congested towns[10] and taxes on parking facilities in cities where through traffic is not a substantial problem. Where the congestion problem is particularly pressing—such as in Central London, Birmingham, Glasgow, etc.—the problem of charging marginal cost might be met by requiring vehicles that use these areas frequently to fit a special milometer with a 'flag' similar to those employed by taxis; and casual traffic to these areas would have to purchase special stickers. The authority would require the flag to be up during the peak hours when the difference between private and social cost is very large.

A complete examination of the administrative problems of a particular plan is far beyond the scope of this paper. In any case much more empirical evidence is required before even informed guesses can

be made. All we can do here is to show in what direction (and how badly) the existing level of taxation is inadequate.

Existing taxes are very much too low for the great majority of urban traffic, slightly too high for rural trunk roads, and much too high for rural subsidiary roads. If it is administratively impossible to levy taxes at different levels in urban areas, and so one set of national motor taxes has to be devised, should they be higher or lower than existing levels? The apparent preponderance of traffic with high cost elasticities certainly suggests that fuel taxes should not be lowered. On the contrary there seems to be a case, albeit inconclusive, for increasing the level of petrol tax. Various other measures might also be considered, such as imposing relatively high purchase taxes or licence duties upon vehicles with a low power-weight ratio; these are the slow-moving vehicles which have low petrol consumption but slow up traffic considerably.

Further Research

The basis of the empirical results in this paper is very slender. Much more data on flows and speeds are required before any extensive blueprint can be devised. However, the collection of suitable data should not be prohibitively expensive. It is much easier to obtain information about flows and speeds than details of numbers of pedestrians and parked cars. One might then enquire whether the estimate of the elasticity from the simple two-variable regression (speed on traffic volume) would give a good approximation to the 'partial' or 'total' marginal cost. From the data for Slough High Street we get an estimate of the elasticity from the two-variable model of 0·50 for average traffic flow. This compares with the 'partial' elasticity in the five-variable model of 0·42. The closeness of this approximation is encouraging—but not too much confidence can be placed in it at this stage.

Acknowledgements

I am grateful to the Road Research Laboratory for supplying much of the data used in this paper. The estimates in this article were considerably improved after correspondence with G. J. Roth. I am solely responsible for any remaining errors.

REFERENCES

[1] Pigou, A. C. (1920). *Wealth and Welfare* (1st edn.). MacMillan, London.

[2] Beckman, M., McGuire, C. B. and Winsten, C. B. (1955). *Studies in the Economics of Transportation*, Yale University Press. (Distributed in Great Britain by Oxford University Press.)

[3] Walters, A. A. (1959). 'The theory and measurement of private and social costs of highway congestion', Discussion Paper B.I. of the Faculty of Commerce and Social Science, Univ. of Birmingham. (An adapted version appeared in *Econometrica*, **29**, 676, 1961.)

[4] Inwood, J. and Green, H. (1954). 'The comparative effects of unilateral waiting and no-waiting restrictions on the speed of traffic in Slough High Street', Road Research Laboratory Note No. RN/2126.

[5] Roth, G. J. (1958). 'The economic benefits to be obtained by road improvements with special reference to vehicle operating costs', Road Research Laboratory Note No. RN/3426.

[6] Smeed, R. J. (1958). 'Theoretical studies and operational research on traffic and traffic congestion', *Bull. Int. Statist. Inst.*, **36**, 347.

[7] Coburn, T. M. (1956). 'Speed on 24 ft. and 30 ft. wide sections of A.5', Road Research Laboratory Note No. RN/2905.

[8] Tanner, J. C. (1957). 'The sampling of road traffic', *Applied Statistics*, **6**, 161.

[9] Department of Scientific and Industrial Research (1960). *Road Research 1959*, HMSO, London.

[10] Walters, A. A. (1954). 'Track costs and motor taxation', *Journal of Industrial Economics*, **2**, 135.

ON THE PRICE CONSCIOUSNESS OF CONSUMERS

ANDRÉ GABOR and C. W. J. GRANGER
The University of Nottingham

The authors report on an enquiry they carried out in Nottingham, to determine the extent to which housewives are conscious of the price of an article at the time of its purchase. They find that for many of the fifteen grocery products covered, a high proportion of housewives were aware of the prices, although the proportion varied somewhat between the products. The results of the enquiry are analysed in several ways and their implications discussed.

Introduction

For over a century economic theory has been under fire because, supposedly, much of its reasoning rests on the postulate that human beings are 'economic men', that is to say, that they are creatures moved solely by selfish material motives.

Not all of this criticism was unfounded. Albeit the name 'economic man' appears to owe its existence not so much to economists as rather to some of their more ironically minded critics, the concept itself was certainly central to classical thought and can be detected in the writings of most representatives of that school. It seems that although the classical economists were well aware of the fact that the world was populated by men of flesh and blood rather than self-centred robots, the strictures upon their teachings were not entirely unjustified, because they failed to distinguish clearly between the two branches of their subject, one of which we now call pure theory and the other descriptive economics.

Traces of this early confusion still linger, although *homo economicus*, explicitly disclaimed by Marshall,[1] has politely surrendered his place to the housewife in the writings of Wicksteed[2] and in many of the later economic textbooks. This substitution of a species of 'economic woman' for 'economic man' meant more than a change in name because, unlike the latter, the housewife was not supposed to be intent on material gain only, nor was utter selfishness attributed to her. But, and here is the rub, she has been represented as a person who will carefully watch prices and adjust her purchases until the pattern of the household budget results in a blissful state of equi-marginal utility.

Since in contemporary Britain, as also in a number of other countries, the housewife is the main administrator of the current expenditure of most families, it would seem to be of some importance to establish whether demand theory provides an acceptable description of her behaviour.

The first question which arises in this connection concerns the extent

to which housewives are aware of the prices of the goods which they regularly purchase. This is the problem which we decided to investigate by means of a sample survey, carried out in November 1958, the method and results of which are discussed in the present paper.

Fundamental Considerations

The assumption which underlies every demand curve is that, all other things being equal, price plays a decisive role in the determination of the rate of purchases of the consumer.

It is well known that this does not apply throughout. The experience of sales organisations operating hire purchase and other similar instalment schemes implies that their prospective customers fix their attention on the required down payment and on the amounts which they have to surrender periodically rather than on the number of instalments, and even to a lesser extent on the total cost of the purchase. There are people who interpret behaviour of this kind as irrational, but we feel that such views are based on a confusion of rationality with a somewhat arbitrary idea of the type of reasoning attributed to 'economic man'. Since rationality means acting in accordance with one's aims, the person whose object is to reconcile the early satisfaction of his desires with the means available to him in the near future can hardly be called irrational if he fails to attach great significance to the sum total of the small periodic payments which he is expected to make in return for an immediate benefit (Gabor and James[3]).

Hire purchase is not by any means the only case where price fades into the background. There are individuals who find it difficult to resist a bargain; if they see something at a temporarily reduced price they will buy it without full regard to the actual cost (and, it is alleged, often without any knowledge of the article purchased). Then, there are situations in which many people feel that it would be inelegant if not invidious to reveal that money is a consideration with them; consequently they will not ask for the price until they have declared that they will buy the article in question. Even if they then find that the price is considerably higher than expected, they will not withdraw for fear of losing face. Again, behaviour of this kind is not necessarily irrational, but in view of the existence of such instances it must be conceded that there are certain transactions to which the customary form of the demand function is inapplicable because the full cost to the consumer is not an important determinant of his decision.

Even if we leave purchases of this type aside, there is still a wide field where traditional demand relationships may be expected to hold sway. Every household makes fairly regularly recurrent purchases of food and other necessities like soap, etc., and, whether it is explicitly stated or not, most theoretical demand curves seem to relate to such commodities because it is here that the basic postulates of demand theory can be satisfied. The housewife has reasonably complete information concerning the needs of the family and she has ample opportunity to

obtain knowledge of the means by which her ends can be attained. But knowledge of the available commodities is not enough to enable us to postulate that her decisions are reached by a maximisation process. Since shopping lists are generally compiled in advance, whether written out or in the form of a mental note, we also have to assume that the housewife is aware of the prices of the commodities in which she is interested.

In view of the relative neglect of cost in the kind of transactions mentioned earlier on, it seemed pertinent to ask whether the significance of price is not being overrated also in the case of recurrent purchases of staple commodities. Occasional questions which we put to housewives revealed that some of them were remarkably ignorant of the prices of those articles which they bought every week. They could not even remember what they paid for butter, flour, eggs, and other regularly bought items, and it appeared that they took notice of price changes only indirectly, via their effect on the purchasing power of their housekeeping money.

This was only a tentative hypothesis, derived from a minute number of instances. We did not expect it to hold throughout and we also realised that the extent to which housewives are conscious of the relevant prices may possibly vary with the income of the household and with the type of shop frequented.

In fact the enquiry which we conducted revealed a much higher degree of price consciousness than we were led to expect. The general implication of the results, set out in more detail later (p. 176), is that there is a fairly wide field where price is of sufficient importance to be remembered. But the fact that housewives are largely cognisant of prices shows merely that cost is likely to have an influence on their shopping lists. The actual strength of this influence and that of price changes is a different question which we intend to explore on a subsequent occasion.

The Sample

(*a*) *How the Sample was chosen.* On the advice of the leading market research firm which assisted us in the planning of the enquiry, a stratified sampling scheme was used, with the numbers so distributed that the higher social classes should be over-represented, thereby allowing the behaviour of these classes to be judged with the appropriate accuracy. The results were later to be weighted, so as to counter this over-representation. This method was adopted because it was clear that in a purely random sample the numbers of persons in the higher social classes would be too small to allow detailed analysis; for instance, in a sample of 500 only about ten persons could be expected to belong to the uppermost of the five social groups which we distinguished.

It was decided to approach 640 housewives at addresses selected from the registers of electors in the City of Nottingham. The new municipal housing estate at Clifton, although technically a part of

Nottingham, was excluded from the scheme, partly because, owing to its geographical position, it was not felt to be an integral part of the city, and partly because the method of stratification adopted failed to give an acceptable indication of the social class of the polling districts of that area. (The reasons for this are given on p. 175-6.)

The choice of the 640 addresses was based almost entirely on the so-called J-index, i.e. on the proportion of potential jurors in the electoral rolls of the appropriate areas. This index was first introduced some nine years ago (Gray, Corlett and Jones[4]), and has since been used as a tool of stratification in a number of social surveys.

In order to obtain sub-areas with roughly comparable population numbers, the smaller polling districts were merged with adjacent ones. This gave us 64 divisions, comprising a total of 109 polling districts. The J-index of each division was then determined and listed in ascending order of magnitude. The list so obtained was divided into three approximately equal strata, and two divisions were chosen at random from the lowest stratum, three from the middle, and five from the upper stratum.

In each of the ten divisions so chosen, the J-index for all the individual 'streets' was then found, a 'street' being a cluster of houses as adjacent as possible and between 80 and 120 in number. The 'streets' of each division were arranged into three groups according to whether they had a low, medium, or high J-index.

With the aid of random numbers, two 'streets' were selected from each of the two lower J-groups and four from the highest J-group. Finally, addresses were drawn at random from each street: six each from those in the lowest J-group, ten each from the medium group, and eight from the highest.

This yielded 64 addresses per division, and the fieldworkers were instructed to attempt to interview one housewife at each of the 640 addresses so chosen. 'Housewife' was to be interpreted broadly as the person responsible for current purchases of provisions for the household.

(*b*) *The Questionnaire.* This consisted of three duplicated sheets with questions about recent purchases of fifteen selected commodities, and a final sheet containing questions concerning the household. The order of the first three sheets was not the same in all the questionnaires, to counteract any possible bias due to fading of interest or growing impatience on the part of the person interviewed.

The fifteen commodities were bacon, tea, coffee, jam and marmalade, breakfast cereal, sausages, potatoes, cheese, household soap, washing powders and detergents, flour, sugar, eggs, butter, and margarine. In the case of each the subject was asked when she bought it last and the fieldworkers were instructed not to press for further information if the last purchase was more than a week ago. In other words, we wanted answers only from those who currently bought the article, because even a person who pays great attention to the price close to the time of the purchase might forget it later on, and seven days were suggested

as the practical limit by the experts of the market research firm who advised us.

Questions were set about the brand or type of the commodity, the quantity bought and the price paid for it, and, finally, about the kind of shop (co-operative shop, grocer, etc.) where the purchase was made.

After the interviewer had gone through the list, the questions on the last page were put to the subject. So as to have a check on the proportions of the sample, we asked about newspaper reading habits, television, household composition, and the occupation and industry of the head of the household. The last question was about income; we did not hope for many answers and indeed the proportion of refusals was so high that the results could not be used in the final analysis.

At the bottom of the page the fieldworker had to enter the date and time of the interview, the name, address, and occupation of the subject, and the interviewer's estimate of age and social group.

The recognisable characteristics of the five social groups listed below were described in further detail in the printed instructions issued to the fieldworkers.

Social Group	Expected Approximate Percentage
A: The well-to-do	2
B: The professional middle class	5
C: The lower middle class	25
D: The working class	60
E: The poor	8

To establish the best way of approach, we first selected a small number of students with some spare-time experience in this type of fieldwork who then made a pilot survey of fifty addresses. The results were reassuring and gave a fair indication of the rate of successful interviews eventually achieved. It was also found that it was advisable to vary the method of approach with the apparent social class of the household: to use the name of the housewife (which was given to the interviewer with the address) only in the case of the lower classes, and to mention that he himself was a student, enquiring on behalf of the Department of Economics of the University, only to subjects who appeared to belong to one of the higher social groups.

(*c*) *The Interviewing Method.* Forty-four students of the University volunteered to do the interviewing, and the addresses were distributed to them on a roughly equal basis. They were carefully instructed, both by word and by printed directions. Substitution of addresses was not allowed and up to three call-backs were required if the housewife was not found at home at the first visit.

It was expected, with the type of questions to be asked and with mostly inexperienced interviewers, that there would be a fairly high number of refusals. This, in fact, proved to be the case, yet the results could not be called unsatisfactory.

The number of successfully completed interviews was 428, or 66·9

per cent of the original number of addresses issued, which is not much below the 70·6 per cent obtained by Durbin and Stuart[5] in a larger survey carried out with the experienced fieldworkers of four professional organisations. Of the 428 completed returns only three were found defective and had to be excluded from the detailed analysis.

(*d*) *The Resulting Sample.* Although the list of addresses had been carefully chosen so as to over-represent the higher social groups, no indication of any such bias was found in the achieved sample. There are several possible explanations of this phenomenon; our impression is that it was largely due to the inefficiency of the J-index.

The social group percentages actually obtained compared well with the percentages which ought to have arisen in an entirely random sample. (The description of the social stratification and the country-wide percentages were placed at our disposal by the courtesy of Messrs A. C. Nielsen Company Limited.)

Social Group:	A	B	C	D	E	Total
Country-wide percentages	2	5	25	60	8	100
Achieved sample	2	7	27	55	9	100

The newspaper readership found in the sample compared favourably with the results found by the recent I.P.A. Survey[6] for the Midlands:

	Daily Mirror	*Daily Express*	*Daily Mail*	*Daily Herald*	*Daily Sketch*	*News Chronicle*
I.P.A.	33	21	12	12	12	5
Sample	37	22	11	5	11	3

The comparison is, of course, not exact, as the I.P.A. survey covers all the Midlands, whereas our sample is entirely an urban one. The *Daily Herald* and the *News Chronicle* might be expected to have 'lost' readers to the local *Nottingham Guardian Journal.* Of the newspapers aimed chiefly at the upper classes, only the *Daily Telegraph* appeared to be slightly over-represented, but its percentage was not high for an urban area.

The readership figures of the Sunday newspapers and a more detailed break-down of the social class reading habits displayed similar correspondence with estimates from other sources.

The percentage ownership of television sets, derived from the sample, may also be compared with the percentage of sets in the entire Midlands, kindly communicated to us by Mr R. J. E. Silvey, Head of the Audience Research Department of the British Broadcasting Corporation.

	BBC/ITV	BBC only	None
Midlands	53	17	30
Sample	55	13	32

The conclusion reached from the analysis of the characteristics of the sample is that it is a good approximation to a *random* sample, i.e. the attempt to over-represent the higher social groups has not succeeded. The results of the survey presented below have all been given under the assumption that the sample is socially unbiased, and no weighting of the figures has been used.

The reason for the failure of the elaborate procedure which was adopted to over-represent the higher social classes is not difficult to find. As mentioned above, the J-index proved inefficient in the sense that it gave poor indication of the social stratum of the population of the areas concerned. Though our enquiry was a relatively small one, and hence only tentative conclusions should be drawn from it, we consider it significant that the application of the J-index left no discernible trace on the proportions of the different social groups of the resulting sample.

The qualification for being a juror is to be under sixty years of age and occupying a house rated at £20 a year or more, and this is not a very stringent limitation on the person's social class. The district with the highest J-index (47·1 per cent) was part of the new municipal housing estate at Clifton, where the majority of occupants are probably D-class. Mr P. G. Gray of the Social Survey, Central Office of Information, to whom we communicated this finding, was aware that both the spread of new local authority housing estates and rating revaluation have tended to make the J-index rather less efficient for stratifying purposes. The evidence of our sample suggests that, at the present time, this index is virtually useless for determining the social class of a district.

The smallness of our sample, coupled with the failure of the sample design to overweight the higher social groups, resulted in such small numbers in some of the cells that it would hardly be justified to draw definite conclusions from them. All the same, we decided not to suppress detail for this reason alone, mainly because we looked upon the whole of our investigation as a pilot survey, the purpose of which was to provide pointers rather than final answers to the problems involved. A larger survey, undertaken on a national basis, would be needed to test some of the hypotheses suggested by our data.

Discussion of the Results

(*a*) *Price Consciousness*. We defined the first measure of price consciousness as the percentage of prices remembered in respect of purchases made within the last seven days, irrespective of whether the price named was correct or not.

The adoption of this measure involved two main assumptions, the first of which was that the housewives did not try to please the field-workers by mentioning an arbitrary figure if they did not in fact have any recollection of the price actually paid.

It seems that this assumption was not on the whole unjustified,

because the scatter of incorrect prices, which is discussed later, implies that, with the possible exception of a small percentage, housewives were under the impression that the price which they named was correct.

The second, more fundamental, assumption was that having some recollection of the price paid can be taken as evidence of being consciously concerned with it. We have no positive proof to support our belief that this is in fact so, but the converse is certainly true: if a housewife cannot remember the price paid for a recent purchase, she could not have paid much attention to it at the time.

Altogether 5276 purchases were reported in sufficient detail for the analysis, and in 4325 cases, or 82·0 per cent, the housewife named a price. Table I shows the analysis of the results by social group and place of purchase.

TABLE I

Number of purchases and the percentage of prices named analysed by social group and place of purchase

Social Group	Place of Purchase					
	Co-op.		Grocer, etc.		All Retail Outlets	
	Number of Purchases	Percent of Prices named	Number of Purchases	Percent of Prices named	Number of Purchases	Percent of Prices named
A	0	—	120	63·3	120	63·3
B	32	87·5	331	64·4	363	66·4
C	360	80·3	1163	81·0	1523	80·8
D	1110	85·9	1785	85·3	2895	85·5
E	132	72·0	243	84·8	375	80·3
All groups	1634	83·6	3642	81·2	5276	82·0

The sample was found to be too small to permit the separation of the entries by the place of purchase into more than the two groups shown in the table. 'Grocer, etc.' therefore includes independent grocers and greengrocers, multiple shops, supermarkets, mobile shops, etc., in other words, all retail outlets not owned by the local co-operative society.

Apart from the fact that the overall price consciousness of housewives appears to be considerably higher than we suspected, the most striking indication of the table is the distribution of price consciousness over the social groups. It is shown to be inversely correlated with social status, with the exception of Group E. And, since family income and social status are certainly directly correlated, we may say that as income falls, concern with prices rises. This was not, of course, an unexpected result, except insofar as the rule appears to be valid only if the poor are excluded. They seem to be less concerned with prices than those in Group D, perhaps because they look upon their small purchases as necessities. Though Group E is represented by 375 purchases only, the effect here discussed does not seem to be due to samp-

ling fluctuations, because the average percentage of prices named by members of this group is below the average of the other groups for each of the fifteen commodities. The pattern of their purchases, i.e. their distribution over the fifteen commodities, does not display any substantial difference from that of the complete sample.

There appears to be a slight general tendency for purchasers at Co-operatives to be better informed of prices than the rest. But the tendency is not fully consistent and the overall result is influenced by the relatively high concentration of Group D on Co-operative shops. Table II shows the distribution.

TABLE II

Comparison of percentages of purchases obtained at Co-op. and at Grocer, etc. retail outlets

Social Group	Co-op.	Grocer, etc.
A	0·0	100·0
B	8·8	91·2
C	23·6	76·4
D	38·3	61·7
E	35·2	64·8
All groups	31·0	69·0

Again we note the apparent anomaly in the behaviour of the poor. An explanation of this may be that housewives tend to buy provisions in shops near to their home, and in the case of Group E, a high proportion of which consists of old people (in our sample about two-thirds were over 65), this tendency is probably stronger than on the average. The Nottingham Co-operative Society controls only about one grocery shop in every ten within the boundaries of the city, and it is possible that in a number of cases considerations of distance have outweighed the preference for the Co-operative Society.

The results of another recent survey[7] partly support this argument. Of the 82 per cent of housewives who said that they had a regular grocer, 37 per cent patronised the nearest shop and 45 per cent one which they could reach in less than five minutes. But it was also found that among those who had a regular grocer, preference for Co-operatives increased with age: under 35, 35%; 35-44, 36%; 45-54, 37%; 55 and over, 45%. The latter results do not necessarily contradict our hypothesis. It is, however, evident that final judgement must be postponed until further specific evidence becomes available.

Table III shows the distribution of price consciousness by commodity and place of purchase. The commodities are listed in descending order of the entries in the last column.

We have no definite explanation to offer for the remarkable dispersion of the percentages named, but it is certainly interesting to note that a number of typically branded goods are found near the lower end of the range, though tea is at the very top. It is, of course, true that housewives frequently describe their favourite brand of tea by

TABLE III

Number of purchases and the percentage of prices named analysed by commodity and place of purchase

Commodity	Place of Purchase					
	Co-op.		Grocer, etc.		All Retail Outlets	
	Number of Purchases	Percent of Prices named	Number of Purchases	Percent of Prices named	Number of Purchases	Percent of Prices named
Tea	145	96·6	231	93·5	376	94·7
Eggs	111	96·4	296	92·2	407	93·4
Sausages	37	83·8	244	91·8	281	90·7
Margarine	97	86·6	167	90·4	264	89·0
Potatoes	59	88·1	320	88·8	379	88·7
Coffee	50	94·0	120	85·8	170	88·2
Bacon	89	83·1	271	84·5	360	84·2
Butter	133	84·2	273	83·5	406	83·7
Jam and Marmalade	86	86·0	162	80·2	248	82·3
Cheese	98	76·5	263	81·4	361	80·1
Washing powders and Detergents	166	88·0	356	74·7	522	78·9
Sugar	152	80·9	273	76·6	425	78·1
Soap	118	74·6	211	71·6	329	72·6
Breakfast cereal	143	81·1	200	62·0	343	70·0
Flour	150	64·8	255	61·6	405	62·7
All commodities	1634	83·6	3642	81·2	5276	82·0

the price; they ask for a quarter of 'the one-and-ninepenny' instead of 'Blogg and Co.'s Special Super Small Leaf Blend', and such a custom would fix the price in their memory. But washing powders, soap, breakfast cereals, and flour are unlikely to be described by the price, and nowadays the price seldom appears on the packet; (522 purchases of washing powder and liquid detergents were reported and in all but 5 cases the housewife remembered the brand name). Hence the latter branded goods would be expected to be low in the table, and tea to be atypical of branded goods. It may well be worth while to test this hypothesis and to explore other possible reasons for the variations in price consciousness.

As mentioned above, the overall figures indicate that those who shop at Co-operatives take more notice of the price than the rest. Yet they may not be bargain hunters in the ordinary sense of the word. According to the results of Bird's survey,[7] only 4 per cent of those housewives whose regular grocer was a Co-operative shop mentioned competitive prices as the reason for their choice and 27 per cent of them thought that the Co-operative grocer charged more than others for commodities like bacon, eggs, and cheese. The most important motive appeared to be the dividend (mentioned by 46 per cent) and almost two-thirds of those who considered their Co-operative shop expensive changed their opinion when they were asked to take the dividend into consideration. Of those whose regular grocer was an independent trader or a branch of a multiple firm, 16 per cent believed that the

prices they paid for bacon, etc., were lower than in other shops and 12 per cent were of the opposite opinion.[7]

(*b*) *Knowledge of the Correct Price.* Of the fifteen commodities listed in the questionnaire, eight were sold at such a variety of prices that it was not practicable to check the answers. The remaining seven, however, could be so checked, and the price named by the housewife could be compared with the appropriate entry in a complete list of the prices of all the brands of the goods concerned, partly obtained from the Nottingham Co-operative Society, and partly extracted from *The Grocer* and other sources.

If there was any departure from the price in our list, the answer was classed as incorrect. Perhaps this was too rigorous a principle, and it seems likely that by applying it we eliminated a number of answers which were in fact correct. No allowance was made for the possibility of so-called 'special offers', nor for the fact that some shops do not strictly adhere to the prices recommended by the manufacturers. Besides, shop assistants are not infallible, and the housewife who reported a price which was, say, a halfpenny below the one in our list may have been charged that price by mistake.

The 'don't know' category includes both those who, for some reason, refused to name the price of the purchase and those who admitted that the price named was more or less a guess. We also had to place in this category a small number of cases where a price was stated but the item was not adequately described and was therefore unidentifiable. It follows that our figures almost certainly understate the extent to which housewives remembered the prices they actually paid. But the downward bias so introduced is probably slight and fairly evenly distributed over the cells.

Altogether 422 housewives reported 2231 recent purchases of one or more of the seven commodities the price of which could be checked. They were (1) tea, (2) margarine, (3) coffee, (4) jam and marmalade, (5) sugar, (6) breakfast cereals, and (7) flour. They are here numbered in the order in which they appear in Table III but, as we shall see, a somewhat different order emerges if the commodities are arranged according to the percentage of *correct* prices returned. Table IV contains a summary of the results.

TABLE IV

Number and percentage of purchases of seven commodities for which prices were named correctly or incorrectly, analysed by place of purchase

Price of Purchase	Place of Purchase					
	Co-op.		Grocer, etc.		All Retail Outlets	
	Number	Percentage	Number	Percentage	Number	Percentage
Correct	432	52·5	708	50·3	1140	51·1
Wrong	249	30·3	382	27·1	631	28·3
Don't know	142	17·2	318	22·6	460	20·6
All purchases	823	100·0	1408	100·0	2231	100·0

The great majority of the housewives reported not more than one purchase per commodity group but the questionnaire provided space for additional items under each general heading, and multiple purchases were in fact returned by a relatively small number. We decided to improve the homogeneity of the data by eliminating all such additional items. The principle adopted introduced a slight upward bias, for where the price of at least one item had been named correctly, we retained one with the correct price. Table V shows the data after elimination of the 343 additional purchases.

TABLE V

Number and percentage of purchases of seven commodities for which prices were named correctly or incorrectly (not more than one purchase per housewife), analysed by social group and place of purchase

Price of Purchase	Place of Purchase					
	Co-op.		Grocer, etc.		All Retail Outlets	
	Number	Percentage	Number	Percentage	Number	Percentage
All social groups						
Correct	413	59·4	663	55·6	1076	57·0
Wrong	184	26·5	296	24·8	480	25·4
Don't know	98	14·1	234	19·6	332	17·6
All purchases	695	100·0	1193	100·0	1888	100·0
Social group A						
Correct	—	—	19	45·2	19	45·2
Wrong	—	—	7	16·7	7	16·7
Don't know	—	—	16	38·1	16	38·1
All purchases	—	—	42	100·0	42	100·0
Social group B						
Correct	7	63·6	47	44·8	54	46·6
Wrong	4	36·4	16	15·2	20	17·2
Don't know	0	0·0	42	40·0	42	36·2
All purchases	11	100·0	105	100·0	116	100·0
Social group C						
Correct	85	57·5	224	56·6	309	56·8
Wrong	40	27·0	86	21·7	126	23·2
Don't know	23	15·5	86	21·7	109	20·0
All purchases	148	100·0	396	100·0	544	100·0
Social group D						
Correct	296	60·9	320	56·6	616	58·6
Wrong	126	25·9	166	29·3	292	27·7
Don't know	64	13·2	80	14·1	144	13·7
All purchases	486	100·0	566	100·0	1052	100·0
Social group E						
Correct	25	50·0	53	63·1	78	58·2
Wrong	14	28·0	21	25·0	35	26·1
Don't know	11	22·0	10	11·9	21	15·7
All purchases	50	100·0	84	100·0	134	100·0

We can now compare the corresponding entries in Tables I, IV and V. Of the 5276 purchases of all the fifteen commodities a price was

named in respect of 82·0 per cent, but taking the seven price-checked commodities we find 79·4 per cent only. After elimination of the 343 additional purchases, the upward bias just mentioned raised the percentage of prices named to 82·4 per cent.

In spite of the very small numbers in some of the cells, Table V displays the same general tendencies as Table I: the percentage of prices named increases as we follow the social scale downwards until we come to Group E, and, again with the exception of Group E, Co-operative customers named higher percentages of prices than those who bought at other shops. The percentages of correct prices returned show a similar picture. However, if we express correct prices as percentages of prices named, a somewhat puzzling pattern arises (Table VI).

TABLE VI

Correct prices as percentages of prices named for 1888 *purchases*

Social Group	Place of Purchase		
	Co-op.	Grocer, etc.	All Retail Outlets
A	—	73·1	73·1
B	63·6	74·6	73·0
C	68·0	72·3	71·0
D	70·1	65·8	67·8
E	64·1	71·6	69·0
All groups	69·2	69·1	69·2

Group D, which is the only group represented by substantial numbers, shows a higher percentage for Co-operative purchases than for others, but its average of 67·8 is the lowest of all. The other social groups are better in respect of their purchases at independent and multiple grocers, and, contrary to what one would expect on the basis of Tables I and V, the last column of Table VI shows decreasing figures down to Group E, which is, on this comparison, slightly better than Group D. It is quite possible that these curious features of Table VI are due to the inadequacies of our data and it seems safest to refrain from drawing any conclusions from them, except that the relationship between prices returned and prices found correct appears to deserve further investigation.

Table VII contains the analysis by commodities, arranged in descending order of the percentage of correct prices named. Generally speaking, the observations appended to Table III apply also here, though the order which emerges is slightly different. Surprisingly high are the proportions of wrong prices in the case of margarine, flour, and breakfast cereals, all three of which are branded goods.

One obvious reason why housewives could recall the price of tea so much better than the price they paid for breakfast cereals is that, generally, only one brand of tea is being purchased by each household whereas, especially in the case of larger families, several kinds of breakfast cereals may be bought, simultaneously or in rotation.

TABLE VII

Number and percentage of purchases for which prices were named correctly or incorrectly, analysed by commodity (1888 purchases) and place of purchase

Commodity	Price of Purchase	Place of Purchase					
		Co-op.		Grocer, etc.		All Retail Outlets	
		Number	Per-centage	Number	Per-centage	Number	Per-centage
Tea	Correct	109	79·6	174	79·1	283	79·3
	Wrong	24	17·5	32	14·5	56	15·7
	Don't know	4	2·9	14	6·4	18	5·0
	All purchases	137	100·0	220	100·0	357	100·0
Coffee	Correct	35	71·5	74	66·7	109	68·1
	Wrong	11	22·4	24	21·6	35	21·9
	Don't know	3	6·1	13	11·7	16	10·0
	All purchases	49	100·0	111	100·0	160	100·0
Sugar	Correct	97	66·9	169	67·1	266	67·0
	Wrong	22	15·2	31	12·3	53	13·4
	Don't know	26	17·9	52	20·6	78	19·6
	All purchases	145	100·0	252	100·0	397	100·0
Jam and Marmalade	Correct	46	63·0	71	57·2	117	59·4
	Wrong	18	24·7	28	22·6	46	23·3
	Don't know	9	12·3	25	20·2	34	17·3
	All purchases	73	100·0	124	100·0	197	100·0
Margarine	Correct	42	46·7	74	45·6	116	46·1
	Wrong	38	42·2	72	44·5	110	43·6
	Don't know	10	11·1	16	9·9	26	10·3
	All purchases	90	100·0	162	100·0	252	100·0
Flour	Correct	41	39·8	59	33·1	100	35·6
	Wrong	29	28·2	53	29·8	82	29·2
	Don't know	33	32·0	66	37·1	99	35·2
	All purchases	103	100·0	178	100·0	281	100·0
Breakfast cereal	Correct	43	43·9	42	28·8	85	34·8
	Wrong	42	42·8	56	38·3	98	40·2
	Don't know	13	13·3	48	32·9	61	25·0
	All purchases	98	100·0	146	100·0	244	100·0

Furthermore, tea is supplied in packages of definite size, mostly $\frac{1}{4}$ lb. and $\frac{1}{2}$ lb., unlike breakfast cereals, which come in a variety of sizes. Also, many individual brands of the latter are available both in small and large cartons at appropriately differing prices, which tends to make price recall still more difficult.

In Table VIII the *percentage correct* figures express the ratio of the number of correct prices named to the number of items bought, i.e., they are averages weighted by the numbers of purchases, whereas the *average percentage scores* are weighted by the numbers of housewives concerned. The latter is the more appropriate measure if we consider the number of items bought by each housewife irrelevant in assessing her

awareness of correct prices. An example will make the difference between the two measures clear. Say a group consists of two housewives, one of whom bought 4 items and named 3 correct prices, while the other bought 6 items and could not state any of the prices correctly. Then percentage correct is $100(3+0)/(4+6)=30\%$, whereas the average percentage score is $(75+0)/2=37{\cdot}5\%$. Had they both bought the same number of items, the two results would agree.

TABLE VIII

Comparison of percentage of prices named correctly with average percentage scores

Social Group	Number of Housewives	Number of Purchases	Number of Prices named Correctly	Percentage Correct	Average Percentage Score
A	9	42	19	45·2	44·1
B	28	116	54	46·6	49·3
C	118	544	309	56·8	57·5
D	229	1052	616	58·6	59·2
E	38	134	78	58·2	60·9
All groups	422	1888	1076	57·0	57·9

Although on the evidence of the percentage correct figures Group E is inferior to Group D, the average percentage scores show a monotonic development throughout. This should be taken into account in assessing the tentative conclusions drawn above in respect of the apparently anomalous behaviour of Group E.

As can be seen from Table IX, the percentage correct figures (which are, of course, since multiple purchases within a group have been eliminated, identical with the percentage scores for each group of items bought) are negatively correlated with the numbers of housewives and the numbers of purchases. This explains why both the average percentage score and the average percentage correct are smaller than the straight average.

It seems that the reason for this phenomenon is the tendency of percentage correct to decline as the number of items bought increases.

TABLE IX

Percentage of prices named correctly according to the number of items purchased

Number of Items Purchased	Number of Housewives	Number of Purchases	Number of Prices named Correctly	Percentage Correct
1	9	9	6	66·7
2	28	56	38	67·9
3	69	207	128	61·8
4	102	408	219	53·7
5	102	510	305	59·8
6	86	516	281	54·5
7	26	182	99	54·4
Total	422	1888	1076	57·0
Average percentage score				57·9
Straight average of the seven percentages				59·8

TABLE X

Percentage of prices named correctly according to the number of items purchased and social group

Social Group		Number of Items Purchased							Total
		1	2	3	4	5	6	7	
A	Number of housewives	0	0	0	6	1	1	1	9
	Average percentage score	—	—	—	41·7	40·0	50·0	57·1	44·1
B	Number of housewives	2	2	4	8	7	4	1	28
	Average percentage score	50·0	100·0	83·3	15·6	60·0	29·2	85·7	49·3
C	Number of housewives	2	5	21	29	24	26	11	118
	Average percentage score	50·0	80·0	58·7	54·3	62·5	52·6	55·8	57·5
D	Number of housewives	3	9	34	57	65	51	10	229
	Average percentage score	66·7	66·7	60·8	59·6	59·7	56·5	52·9	59·2
E	Number of housewives	2	12	10	2	5	4	3	38
	Average percentage score	100·0	58·3	63·3	62·5	52·0	66·7	42·9	60·9
All social groups:	Number of housewives	9	28	69	102	102	86	26	422
	Average percentage score	66·7	67·9	61·8	53·7	59·8	54·5	54·4	57·9

Why exactly this should be so was not revealed by our enquiry. One possible explanation is that the larger the number of items recently bought, the less reliable is the housewife's recollection of the prices, assuming that the total number of goods recently bought by the housewife varied directly with the proportion which she bought of the seven price-checked commodities. Another is that after the first few questions some of the housewives may have realised that they would be asked the price of each commodity reported as recently purchased and preferred to keep silent about those items the prices of which they could not well remember.

Table X shows the results in further detail. As in Table IX, percentage correct and percentage score have the same value for each of the cells but not for the totals shown in the last column of Table X.

The modal numbers of items bought are 4 and 5, with the exception of Group E. In spite of the small numbers in the cells, the indication that the poor buy less items in the course of an average week is probably correct. One likely reason is that their households tend to be smaller than those of the other four groups.

(*c*) *Deviations from the Correct Price.* As shown in the preceding discussion, of the 1888 purchases of the seven price-checked commodities 1076 or 57·0 per cent were correctly priced by the housewives. The number of incorrect prices named was 480, hence it may be said that an additional 25·4 per cent *believed* they knew the price.

In order to get some impression of the magnitude and distribution of the deviations from the correct price, we subjected the purchases from Co-operative shops to a detailed analysis. Our main reason for selecting this category was the belief that the prices obtained from the Nottingham Co-operative Society were probably more reliable than those which we extracted from other sources. In fact this may not have been the case for, as we were later informed, individual branches of the society do not invariably adhere to the centrally compiled price list. The results are summarised in Table XI.

TABLE XI

Frequency distribution of 184 *wrongly named prices of seven commodities purchased from Co-operative shops*

Percentage Deviations from Correct Price		Number	Per Cent	
Exceeding	Not Exceeding			
+50	+60	1	0·5	
+40	+50	0	0·0	
+30	+40	2	1·1	40·2
+20	+30	5	2·7	
+10	+20	18	9·8	
0	+10	48	26·1	
0	−10	47	25·5	
−10	−20	34	18·5	48·4
−20	−30	6	3·3	
−30	−40	2	1·1	
Not classifiable		21	11·4	
Total		184	100·0	

Of the 95 prices in the two central groups 41, that is to say less than half, differed from the correct price by not more than 5 per cent: 20 in the positive direction and 21 in the negative direction. This apparent irregularity is almost certainly due to the fact that many of the errors recorded were either a penny or a halfpenny, and a penny error on a unit price which is below 1s. 8d., as also a halfpenny error if the actual price is less than 10d., amount to deviations of over 5 per cent. Otherwise the distribution slopes away in both directions and is as symmetrical as could be expected with such small numbers. The analysis by social group and by commodity has not revealed any noteworthy features.

As shown in Table V, of the 695 purchases from Co-operative shops 413 prices or 59·4 per cent were correctly named. Had we adopted the principle of accepting a price as correct if the deviation did not exceed 5 per cent, this would have raised the percentage correct to 65·3, and a ± 10 per cent margin would have raised it to 73·1.

The 21 prices which were not classifiable were obtained from housewives who failed to name the actual brand or type of the commodity and who stated a price which did not tally with any item under the appropriate heading in the full list obtained from the Nottingham Co-operative Society.

Concluding Remarks

There can be little doubt that the files of commercial market research organisations contain a great deal of data relevant to the purpose of this study, but since we have not found it possible to gain access to such material we had to undertake our own enquiry. We think that it was well worth while because, in spite of the smallness of the sample, evidence has been produced which shows that whether housewives are 'economic women' or not, a very substantial proportion of them take cognisance of current prices and can therefore behave in a way which conforms to the basic postulates of demand theory.

In addition, the distributions obtained exhibit a number of interesting features. Some of these may well be due to the inadequacy of our sample and hence all conclusions drawn from them must be tentative. We feel that the enquiry should be repeated on a sufficiently large scale to test the hypotheses put forward and to distinguish also between purchasers from individual grocers, self-service and multiple shops which display all prices, market stalls, etc., and to assess price consciousness according to certain other characteristics which we had to neglect, such as cash and credit purchases, personal shopping, and telephone orders. It would be even more valuable to establish the extent to which price changes affect the rate of purchases of different social groups.

Acknowledgements

Our thanks are due to the University of Nottingham for providing us with the finance and facilities for carrying out the enquiry, to the

students of the University who co-operated in the field work, to Miss Maureen Sherwin who carried out the statistical analysis of the data and, last but not least, to Messrs A. C. Nielsen Company Limited for most valuable advice, freely and generously given, in connection with the design of the questionnaire and the sampling method.

REFERENCES

[1] Marshall, A. (1890). *Principles of Economics*, (1st edition), p. 78. Macmillan, London.

[2] Wicksteed, P. (1910). *Common Sense of Political Economy*, p. 18 and *passim*. Macmillan, London.

[3] Gabor, A. and James, S. F. (1958). 'On measuring utility', *Yorks. Bull. of Econ. and Soc. Res.*, **10**, 1.

[4] Gray, P. G., Corlett, T. and Jones, P. (1951). *The Proportion of Jurors as an Index of the Economic Status of a District.* Pamphlet of the Social Survey, Central Office of Information, London.

[5] Durbin, J. and Stuart, A. (1954). 'Callbacks and clustering in sample surveys; an experimental study'. *J. R. Statist. Soc.*, A, **117**, 387.

[6] Institute of Practitioners in Advertising. *The National Readership Survey*, London.

[7] *Mrs Housewife and Her Grocer*, pp. 11 and 14. Alfred Bird and Sons, Ltd., Devonshire Works, Birmingham 12. (Published for private circulation only. Price 15s.)

BOOK REVIEWS

M. S. Bartlett: *Stochastic Population Models in Ecology and Epidemiology*. London: Methuen, 1960. Pp. x+90. 12s. 6d.

This is the fifth in Methuen's series of monographs on applied probability and statistics and is written by the general editor of the series. The book opens with an account of the way in which Poisson, negative binomial, and logarithmic distributions can be generated in ecological situations. The next three chapters deal with the stochastic analogues of the classical deterministic models of population growth, prey-predator interaction, and competition between species. The book closes with an account of stochastic models for epidemics.

The author states in the preface that 'this book will be most useful to biometricians with some acquaintance already with stochastic (probabilistic) population models and recent allied statistical developments'. The mathematical style is highly condensed, and presupposes considerable technical facility (particularly in the use of probability generating functions). It seems unlikely that the author's hope will be fulfilled that the book will also serve as an introduction to the subject to theoretically-minded ecologists. However, those who work in the field of stochastic processes will find a wealth of information and of new ideas and methods; of particular interest are the methods of finding approximate solutions of intractable equations.

The author shows a healthy interest in finding practical realisations of the theoretical models; his own work on the critical population size which determines whether measles can maintain itself in a community without re-introduction from outside is of particular importance. However, it is rather disappointing that the only realisations of the ecological prey-predator and competition models are the rather artificial laboratory experiments of Gause and Park. It remains to be seen how valuable stochastic models will be in natural ecological situations. M. G. BULMER

K. A. Brownlee: *Statistical Theory and Methodology in Science and Engineering*. New York and London: John Wiley and Sons, 1960. Pp. xv+570. 134s.

This textbook arose from a university course. In the words of the author this course was 'partly for students of the experimental sciences and partly for statistics majors, the main objective being to give both groups some facility and self-confidence in the actual use of statistical methods'. The book is a blend of theory and practice, as its title implies; the author has taken great pains to show not only how a method works, but why it works.

The reader should be familiar with elementary college algebra (roughly A-level mathematics); a knowledge of elementary calculus is useful but not essential. Some of the proofs are quite lengthy and consequently the algebra becomes somewhat tedious. Hence I feel that this is a book which is much better studied as part of a course, in which a certain amount of hard work is to be expected, than tackled as casual reading by the scientific worker. The latter reader is tempted to skip the proofs and turn to the results, thereby vitiating one of the purposes of the book. Yet even he may, by

judicious reading, be enabled to gain an insight into the principles underlying some of the techniques that he applies. I sometimes felt that the author's earlier booklet *Industrial Experimentation* was so avowedly a 'cookbook' that it might be dangerous to the uncritical reader. It is interesting to see in the present work what Professor Brownlee can do when he chooses to give both the recipes and their backgrounds; one should not grumble if the going is, at times, rather heavy.

It would be tedious to enumerate the contents; some idea of the level attained may be given by stating that the coverage extends to multiple regression, four-way analysis of variance, and simple experimental designs. I found the discussion of the various models used in the analysis of variance very helpful. It is sufficiently general to enable the reader to tackle some of his problems that do not conform with any of the straightforward textbook categories.

Professor Brownlee does not enter the confidence-fiducial limit controversy, but usually keeps to confidence limits. However, on occasion he claims more for these limits than they are capable of; I refer to the confidence limits for the binomial and Poisson expectations. Here Brownlee gives limits as, for example, 95% limits when all that can be said, on the strict confidence limit basis, is that the level of confidence is at least 95%.

As a matter of personal taste I was sorry that, amid the quite complex cases treated under the heading of regression, some space was not found for functional relationships. They are often of interest to scientists.

One feature which I found somewhat disconcerting was the use of $\chi^2(n-1)$ to signify χ^2 with $n-1$ degrees of freedom. This is done with no variation of size of type or use of subscripts; I expect that one soon becomes accustomed to the convention but at first the interpretation of complicated expressions containing χ^2 demands extra care on the part of the reader. I would protest strongly at the use of expressions such as $t(rt(n-1))$, in which the first t is Student's t, and the second t is the number of columns in a two-way analysis of variance, the whole expression denoting Student's t with $rt(n-1)$ degrees of freedom. I think a more imaginative use of symbols is called for. I have not attempted to check the algebra and it would be surprising if, in a book of this size, an occasional misprint did not creep in. The only one I noticed was on page 407, where in equation (10.4) the last subscript should be η and not λ.

This is an expensive book; it may be recommended to scientists and engineers who like to know the why and wherefore of the statistical techniques they wish to use, but who lack the mathematical equipment to tackle the more advanced treatises.

D. G. BEECH

W. Edwards Deming: *Sample Design in Business Research*. London and New York: John Wiley and Sons, 1960. Pp. xx+517. 96s.

A new book on sampling by such a distinguished exponent of the science as Professor Deming naturally arouses the liveliest expectations: and, on the whole, they are satisfied. He has produced a substantial compendium, described in his preface as a text-book, covering virtually the whole field of sampling, both practice and theory. It first lays down a rigorous code of standards for statistical consultants called in to advise on sampling and then instructs them how to do the job so that these rigorous standards are fully met. The third part consists of an economical but thorough presentation of

the theory of sampling. Throughout the book the author concentrates largely on interpenetrating replicate designs, stressing their value in facilitating standard error computations. It is extensively illustrated with examples drawn from a wide variety of sampling enquiries and is by no means confined to sampling business records such as accounts or stores inventories. For a detailed description of the book, including a careful listing of errors (in fact they seem to be comparatively few in number), reference should be made to the lengthy review by Kish in the *Journal of the American Statistical Association*, Vol. **56**, No. 295, pp. 741-4.

The book does not replace the excellent sampling text-books already available, nor was it intended to. This being so, the question may be asked: 'Is there room for yet another book on sampling?' The answer must surely be that there is at least room for *this* book. Few who are engaged in sampling can fail to profit from it. Some readers may quarrel with the emphasis laid on different aspects of sampling procedure. Stratification is dealt with only at a fairly late stage, and the advantages to be gained by intelligent stratification seem to be inadequately stressed. On this side of the Atlantic it may be felt that some of the examples are not of much immediate value. For instance, there is a detailed chapter on the interviewers' task in enumerating areas and then selecting 'segments' at random. This is not the sort of problem many of us face here. Such examples, however, are none-theless valuable in showing how practical problems have been tackled without in any way lowering statistical standards.

It is not altogether an easy book to read or find one's way about in. Certain topics, such as non-response, are dealt with in a number of different paragraphs scattered throughout the book. But on the whole the experienced practitioner in sampling will find in it a great deal of sound practical advice and much thought-provoking comment. B. P. EMMETT

Gerald J. Lieberman and Donald B. Owen: *Tables of the Hypergeometric Probability Distribution.* California: Stanford University Press; London: Oxford University Press, 1961. Pp. vi+726. £6.

The hypergeometric probability distribution is not easy to compute because of the laborious calculations with factorials that are entailed. Consequently in most problems in which it is involved it has usually been approximated by means of binomial, Poisson, and normal distributions. The present volume is the first tabulation of the distribution on an extensive scale—a fruit of what may be called the electronic era in table compilation. The simplest application of the distribution is in the following situation:

If there is a population of size N, of which k individuals have a certain characteristic, and a sample is drawn of size n, then the probability that the sample contains x individuals with this characteristic is given by the hypergeometric frequency function $\mathrm{p}(N, n, k, x)$. The probability of finding x or fewer individuals with this characteristic is given by the cumulative function $\mathrm{P}(N, n, k, x)$. Since the individual and cumulative probabilities are unaltered when n and k are interchanged, it is necessary to tabulate only for $k \leqslant n$. For both individual and cumulative probabilities the amount of tabulation has been considerably reduced by taking into account three additional symmetries. Consequently these symmetries may have to be applied when the tables are used.

The probabilities are given to six places of decimals. In the following description of the tables it should be remembered that always $k \leqslant n$. The

tables give all possible hypergeometric distributions for $N \leqslant 25$. For $25 < N \leqslant 50$ all possible distributions are given except that one of the set of three symmetries may have to be used. For $N = 60(10)100$ all values are given except that one of the three symmetries may have to be used. For $N = 1000$, $n = 500$, all distributions are given except that probabilities equal to 0 to six decimal places are not explicitly given and that one of the three symmetries may have to be used. A table of probabilities is also given for $N = 100(100)2000$, $n = \frac{1}{2}N$, $k = n - 1$, n, and all values of x. In addition, there is a table of log factorial N to fifteen decimal places for $N = 1$ to 2000.

The introduction gives various applications of the tables: a sequential procedure (I found the explanation a little difficult to follow), 2×2 contingency tables, distribution of the number of exceedances, Bayesian prediction for the binomial distribution, and sampling inspection. Other sections in the introduction deal with various approximations to the hypergeometric distribution, interpolation in the tables, and some useful formulae on sums of combinatorials. A bibliography is given.

The quality of the photographic reproduction is generally quite adequate, though here and there individual figures have suffered. This collection of about 700 pages filled with columns of figures is a remarkable example of the liability of relatively simple problems in probability to need a great amount of tabulation to deal with the numerous contingencies that may arise. The tables will be useful to the specialist in sampling problems, but I fancy that many workers who encounter problems of sampling without replacement will continue to find the accuracy obtainable by the aid of various approximations quite sufficient for their needs. D. G. BEECH

INDEX

Bartholomew, D. J., 52
Bartlett, M. S., 189
Beattie, D. W., 147
Beech, D. G., 14, 64, 112, 190, 192
bias, 'not-at-home', 52
Brownlee, K. A., 189
Bulmer, M. G., 189

continuous production, sampling plan for, 147
Conway, Freda, 9

Danø, S., 112
Deming, W. E., 190
dynamic programming:
  and decision theory, 39
  introduction to, 32

Emmett, B. P., 191
Enrick, N. L., 64
expenditure enquiry:
  inter-week variations, 136
  variability in 1953-54, 117

factorial experiments, confounded, 83, 98
Furneaux, W. D., 113

Gabor, A., 170
Granger, C. W. J., 170

Herdan, G., 65

Inkson, R. H. E., 98
interviewer differences, 93

Jung, A. F., 93

Kemsley, W. F. F., 117

Lieberman, G. J., 191
Lindley, D. V., 39
'Linear Programming in Industry', review, 112
linguistics, Simon's model of distribution functions in, 65
Loveday, R., 112

measurements and standards, 9
'Modern Probability Theory and Its Applications', review, 63
motor taxes, evidence on optimum, 157
multivariate analysis, industrial applications of, 1

Owen, D. B., 191

Parzen, E., 63
price consciousness, 170
Pridmore, W. A., 64

'Quality Control', review, 64

Read, D. R., 147

'Sample Design in Business Research', review, 190
sample surveys, allowing for 'not-at-home' bias in, 52
sampling plan, variable lot-size, 147
Sandon, F., 113, 115
selection, truncated, application of non-parametric statistics to, 77
Simon's model, 65
Simpson, M. G., 32
'Statistical Theory and Methodology in Science and Engineering', review, 189
'Statistics, A Second Course in', review, 112
'Stochastic Population Models in Ecology and Epidemiology', review, 189
straight line, precision of the gradient, 14

'Tables of the Hypergeometric Probability Distribution', review, 191
Tallis, G. M., 77
'The Chosen Few: An Examination of Some Aspects of University Selection in Britain', review, 113
Thomas, B. A. M., 1
truncated selection, non-parametric statistics applied to, 77
Turner, R., 136

Walters, A. A., 157
Wilkie, D., 83